

Coping with peer rejection

Accounts of rejected Nobel-winning discoveries highlight the conservatism in science. Despite their historical misjudgements, journal editors can help, but above all, visionaries will need sheer persistence.

Not many people spend tens of thousands of dollars to tell the world that they were robbed. But that is what Raymond Damadian and his company did last week when he discovered that he hadn't won the Nobel Prize in Physiology or Medicine, and complained in full-page advertisements in *The New York Times* and other prominent newspapers (see page 648). He claims in his advertisement that he should have shared the prize won by Paul Lauterbur and Peter Mansfield for their work on magnetic resonance imaging.

Whatever the merits of Damadian's case, the episode highlights the difficulties in assessing ground-breaking work. If it is controversial 30 years after the event, one can imagine the divergences of opinion that arise over truly innovative research before history has had a chance to consider its verdict — when papers are submitted to journals and applications sent to grant funding panels. In the latter case, researchers can keep their ambitions masked, stating goals that are predictable extensions of previous work, thereby — as cynics would have it — maximizing the chance of funding. But in the case of journals, they have no choice but to stake their genuine claim.

Some funding agencies, to their credit, are setting out to encourage riskier, visionary applications. The US National Institutes of Health has a new roadmap that includes a Director's Challenge with precisely that aim. The European Commission is also setting up a fund for visionary research, the New and Emerging Science and Technology programme, which will shortly issue a call for proposals (see www.cordis.lu/nest/home.html).

Regrets

What of the journals? *Nature*, while proud of its content over the years, has a confession to make about this year's medicine Nobels. Not so long ago, presciently pleased with having published Lauterbur's work, we celebrated it along with other *Nature* greats in a promotional campaign. Lauterbur politely wrote in to point out that we had published it only after he had appealed against a rejection.

In case anybody runs away with the idea that *Nature* is unusually culpable in this respect, they can look at a collection of rejections experienced by Nobel winners that neatly illustrates the hurdles they had to overcome to publish their work. Juan Miguel Campanario, a physicist at the University of Alcalá in Madrid, Spain, has compiled a list of more than 20 Nobel laureates' rejections by many journals, and recollections by many more of resistance by their peers (see www2.uah.es/jmc).

Not all of *Nature's* Nobel-winning casualties are totally embarrassing for us. Our notorious rejection of the Krebs cycle in 1937 is partly mitigated by the fact that we said we would publish it once several weeks' congestion was out of the way, only for Krebs to take it elsewhere. In some cases cited by Campanario, we are accused only of having the nerve to force the authors to shorten their papers.

But there are unarguable *faux pas* in our history. These include the rejection of Cerenkov radiation, Hideki Yukawa's meson, work on photosynthesis by Johann Deisenhofer, Robert Huber and Hartmut Michel, and the initial rejection (but eventual acceptance) of Stephen Hawking's black-hole radiation. Hindsight is always per-

fect. But we can take comfort, however dubious, from the fact that our unmitigated embarrassments are but a minority in a substantial list of journals' historical misjudgements.

We can take more respectable comfort from a little-celebrated positive accomplishment of editors, which is to champion submitted papers in the teeth of referees' (and sometimes colleagues') resistance. One such submission, according to his Nobel lecture, came from Thomas Cech. The three referees ("outraged enzymologists", as Cech described them) all opposed the idea that self-splicing RNA could be a catalyst, but *Nature* published it nevertheless.

Reasons to publish

A straw poll of *Nature* journals' editors confirms that risk-taking and hopefully enlightened acceptance by editors persists — although whether at the Nobel level it is too early to say. Confidentiality prevents us from being specific, but papers in (for example) stem-cell development, cell signalling networks, genetic linkage to disease, telomerase dysfunction and extrasolar planets were accepted for publication in recent years despite significant scepticism, and were subsequently well cited.

Other examples, for instance in mammalian evolution and muscle crossbridge dynamics, were published with editors and referees suspecting that their conclusions were probably wrong but giving the papers the benefit of the doubt because there were no insurmountable technical objections and they seemed important. Such cases have proved stimulating for their fields, even though (in at least one case) the conclusions, as techniques have improved, have indeed required revision.

What are the morals of these tales? Certainly we need a diversity of good journals. The laureates' rejected papers ended up being published somewhere respectable. And in particular there is perhaps some continuing virtue, along with the pitfalls, in the old elitist model of learned-society journals. The *Proceedings of the National Academy of Sciences*, for example, has published innovative papers that had failed to be appreciated by editors elsewhere, because the authors were academy members and so were able to publish by right.

This is strikingly reminiscent of perhaps the most celebrated editorial judgements of all, in *Annalen der Physik* in 1905. That was the year in which Einstein published five extraordinary papers in that journal, including special relativity and the photoelectric effect. The journal had a great editor in Max Planck. He recognized the virtue of publishing such outlandish ideas, but there was also a policy that allowed authors much latitude after their first publication. Indeed, in journals in those days, the burden of proof was generally on the opponents rather than the proponents of new ideas. One might also remember just how exceptional Nobel-winning discoveries tend to be. By and large, peer-filtering has strong virtues.

Nevertheless — a final moral — rejected authors who are convinced of the ground-breaking value of their controversial conclusions should persist. A final rejection on the grounds of questionable significance may mean that one journal has closed its door on you, but that is no reason to be cowed into silence. Remember, as you seek a different home for your work, that you are in wonderful company. ■





Out of the loop
Medical-imaging pioneer disputes Nobel award
p648



Shake-up
Japanese earthquake gives tsunami experts a wealth of data
p649



Past crimes
Germany reassesses its Nazi-influenced scientific past
p650



Lucky streak
Teenager's photo finds fame on NASA website
p653

Health chiefs poised to step up US scrutiny of microbe research

Erika Check, Washington

The United States is set to implement formal structures for dealing with biological research that could aid terrorists, in the light of recommendations issued by the National Academy of Sciences.

An academy panel was charged with advising the federal government on the handling of biological findings that could provide information on engineering dangerous pathogens or weakening treatments against them. The study, supported by two non-profit groups — the New York-based Alfred P. Sloan Foundation and the Nuclear Threat Initiative in Washington DC — began before the terrorist attacks of 11 September 2001. But it acquired new urgency after the anthrax attacks of that autumn, and with a batch of research findings whose open publication concerned some government officials.

The study's report, released on 8 October, recommends that the Department of Health and Human Services create a committee to advise it on issues related to biological research not intended to do harm, but that could conceivably threaten national security. The council also recommends that the National Institutes of Health reconstitute its Recombinant DNA Advisory Committee (RAC) to review such research in cases where local biosafety committees are unsure whether to allow the work to proceed.

The report defines seven categories of research that should always be reviewed in this way. Some are already reviewed by the RAC. But the report cautions strongly against restricting the types of research that scientists can pursue or publish.

"Vigilant self-government is the goal," says Gerald Fink, a molecular biologist at the Massachusetts Institute of Technology who chaired the panel.

Senior health-department officials say they will decide by early November whether to implement the report's recommendations. But Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, says that the National Institutes of Health has already decided to appoint a subcommittee to the RAC for dealing with national-security issues raised by biological research.



Think tank: an expert panel has drafted rules to monitor US labs that culture dangerous pathogens.

The kind of recommendations made by the report have been under discussion in the scientific and national-security communities for several months. Many scientists agree that a RAC-like body would be best placed to deal with issues raised by research in 'areas of concern', such as experiments that might change which animals a pathogen can infect.

Although the White House's Office of Science and Technology Policy had coordinated some of these discussions, its director John Marburger says the health department will be taking the lead on the issue. "We've been involved in some of the discussions so far, but the expertise for human biology lies within the department," he says.

Scientific groups, meanwhile, are relieved that the Fink panel has not proposed any further research restrictions, although they are still wary that its recommendations may be added to during their implementation. "We're going to have to watch as this goes forward," says Janet Shoemaker of the American Society for Microbiology. "It's a broad architecture but it has to be implemented in such a

way that it doesn't inhibit research." Other observers say that the most difficult questions raised by biological research in the light of terrorism are only now starting to emerge.

For instance, they point out, the RAC operates in full public view, but it is unclear whether this will be the case if it starts to deal with security-related issues. And some say that the seven research types of concern in the report do not include those that will raise the most difficult issues. The experiments of concern all relate to microbiology and infectious disease — for instance, how to make a pathogen more infectious or resistant to antibiotics. They do not include other potentially threatening research, such as experiments in manipulating brain function.

"The experiments they've highlighted are the low-hanging fruit," says Gigi Kwik of the Center for Civilian Biodefense Strategies at Johns Hopkins University in Baltimore, Maryland. "We really need to do something about research that falls into the grey area. And filling in those details is going to be the difficult, squishy work ahead."

MAXIMILIAN STOCK/SPL

Routine tests reveal unknown strains of BSE prions

Marika Willerroider, Munich

New strains of the prion that causes bovine spongiform encephalopathy (BSE) have been reported by scientists in Italy and Japan.

Until now, all known cases of BSE have been caused by the same prion strain. But experts have long suspected that additional strains of BSE might be lurking in cattle, which they fear could propagate new human diseases, as happened in the 1990s outbreak of variant Creutzfeldt–Jakob disease.

The new strains are readily detected by existing tests for BSE. The researchers identified them from brain samples using a technique called the western blot, which produces a series of bands that represent the prions present. In the case of the new strains, the western blot showed a different pattern of bands from conventional BSE. Researchers remain concerned that if prion variants continue to emerge, some of them may escape detection by existing tests.

On 6 October, Morikazu Shinagawa of the National Institute of Animal Health in Tsukuba, Japan, reported that a bullock born there had tested positive for BSE, and that the characteristics of the prion were different from those of classical BSE. Two days later, Maria Caramelli at the National Reference Centre for Animal Encephalopathies in Turin, Italy, reported that two of 110 cows that had tested positive for BSE involved similarly distinctive prions.

In both cases, the prions were found during abattoir sampling of symptom-free animals. The Japanese and Italian scientists have yet to confirm whether the two strains are identical.

“What we have is definitively a new variant, but it was detectable by a conventional rapid test, so there is no immediate risk of a new epidemic,” says Caramelli, who presented her work at the International Prion Conference in Munich last week.

“There has never been good reason to believe that there should only be one variant in cattle,” says Adriano Aguzzi, a prion expert at the University Hospital in Zürich, who points out that several strains of spongiform encephalopathy have been found in sheep and in humans. “It will be interesting to find out if the new prion variants were introduced in cattle feed, or if they spontaneously emerged in the cattle as a result of a gene mutation,” he adds. ■

Physician launches public protest over medical Nobel

Helen Pearson

It should have been a time of celebration, but instead the Nobel prizes were last week caught up in a very public row. Angered by his exclusion from this year's medicine award, a New York-based inventor took out adverts in US newspapers calling for a share of the prize.

On 6 October, physicists Paul Lauterbur of the University of Illinois, Urbana-Champaign, and Peter Mansfield of the University of Nottingham, UK, won the Nobel Prize in Physiology or Medicine for their role in developing magnetic resonance imaging (MRI) as a medical technique for scanning the human body (see *Nature* **425**, 547; 2003).

But the announcement triggered a hostile reaction from Raymond Damadian, a physician who claims to have invented the technique. He protested by placing full-page advertisements in last week's *The Washington Post*, *The New York Times* and *Los Angeles Times*. “If I had never been born, there would never be MRI today,” he told *Nature*.

Damadian's advertisements ask readers to cut out a coupon and send it to the Nobel committee demanding that he be recognized as the joint third winner of the prize — a forlorn hope, as Nobel committees never discuss the justification of their awards. Damadian says that he has not yet totted up the full cost of the adverts, which were paid for by his MRI imaging company FONAR in Melville, New York, but a news report in *The Washington Post* valued the slot in that newspaper alone at about \$80,000.

Damadian says that he originally came up with the idea of using the then-established physical scientists' technique of nuclear magnetic resonance (NMR) to distinguish between states of body tissue, an approach that became known as MRI. In 1971, he published a paper in *Science* showing that cancerous and healthy tissue can be told apart by the different NMR signals that they emit (R. Damadian *Science* **171**, 1151–1153; 1971). He later patented a technique to scan the body for tumours by taking a series of NMR readings from different spots.

MRI experts agree that Damadian's concept was important, but say that without Lauterbur and Mansfield's contributions the MRI technique would not be where it is



Visual aid: Raymond Damadian built this magnetic resonance imaging machine, dubbed Indomitable, in 1977.

today. The pair showed how to use a graded magnetic field, which varies from strong to weak across an object, to rapidly take its image. “The real breakthroughs came from Lauterbur and Mansfield,” says Richard Ernst, of the Swiss Federal Institute of Technology, Zurich, who won the 1991 chemistry Nobel for improving the resolution of NMR.

Researchers add that concerns about Damadian's response — which they say was anticipated before the announcement — have held up the award of a Nobel prize for MRI. The Nobel committee, Mansfield and Lauterbur all declined to comment.

Disputes within the scientific community often dog the Nobels, but observers say that this latest protest in the wider media breaks new ground. “It's absolutely sensational. I've never heard of anything like that,” says Tore Frängsmyr of Uppsala University in Sweden, who edits the Nobel Foundation yearbook.

Frängsmyr and others predict that publicized scuffles over the Nobels might become more common in the future, partly because today's top scientists increasingly hold celebrity status. Bruised egos are particularly likely in medicine, where discoveries are often made by large collaborative groups — the rules state that Nobel prizes can be awarded to no more than three individuals (see *Nature* **413**, 560–564; 2001).

Whatever the outcome of Damadian's complaint, observers say that such arguments could diminish the esteem in which the Nobel prizes are held. The prize-giving ritual could come to eclipse the underlying achievement, warns science historian Robert Marc Friedman of the University of Oslo, Norway. “It would be good if there was a lot less hysteria around the prize,” he says. ■

Earthquake makes waves for tsunami models

David Cyranoski, Tokyo

Researchers investigating how tsunamis form are awash with data following an earthquake in Japan late last month.

The quake, which had a magnitude of 8.0, struck off the southeastern shore of the northern island of Hokkaido on 26 September, sending waves up to four metres tall rushing towards the coast. But the quake's epicentre was almost directly beneath an ocean-bottom monitoring system, containing instruments that use pressure gauges to measure rises and dips in the ocean floor.

Scientists are now rushing to plug the data from these instruments into models of tsunami generation, which they say should help them to improve predictions of the waves and perhaps prevent future fatalities.

Tsunamis — waves that can reach up to 30 metres high — are created by sudden jolts to the sea floor. Although instruments have monitored passing tsunamis before, they have never collected data from the spot where such a wave is born, says Kiyoshi Suyehiro, director of the deep-sea-research department at the Japan Marine Science and Technology Center in Yokosuka. These data have been the missing link to understanding tsunami generation, he says. "We now have actual data to test models that incorporate the earthquake faulting, resulting sea-bottom deformation, and responding seawater mass movement," says Suyehiro.



After the storm: Japanese fishing boats languish on the shore following last month's tsunami.

That should help to make tsunami warnings more accurate. Such alerts are already common in earthquake-prone Japan, where seismic waves travelling at 20,000 kilometres per hour are used to predict the arrival of tsunamis, which travel at a mere 750 km per hour. But those predictions are based on a

fairly simple model and are often wrong, says Kenji Satake, an earthquake researcher at the National Institute of Advanced Industrial Science and Technology in Tsukuba, Japan. The tsunami from the September quake, for example, was twice as high as predicted by the Japan Meteorological Agency.

"Basically a big earthquake is taken to mean a big tsunami," says Satake. But that is not always the case. Some tsunamis are caused by mild earthquakes that pack a bigger punch in terms of making waves. This phenomenon isn't fully understood, but is thought to depend on factors such as the direction and speed of the seafloor slip along the fault.

Researchers still need to clean up the data from the recent quake, to adjust for the effect of water temperature on the measurements, for example. "Once that is done, they can be used to understand the dynamics of tsunami generation," says Satake. Suyehiro is already racing to apply the data to his own model. "We're working as fast as we can," he says.

Others, such as Charles Sollitt, an emeritus civil engineer at Oregon State University's Wave Research Laboratory in Corvallis, will use the data to study how far, and how fast, tsunamis can flood over the coastline. But Harry Yeh, a tsunami expert at Oregon State University, notes that progress might be slowed by the fact that a lot of the information posted online about the data is written only in Japanese. ■

Bleak forecast for space weather

Tony Reichhardt, Washington

The main US centre for predicting and monitoring solar storms could be shut down unless government agencies and other users of 'space weather' forecasts rally to its defence, according to scientists and congressional staff.

The Space Environment Center (SEC) of the National Oceanic and Atmospheric Administration (NOAA) in Boulder, Colorado, has had its \$8.3-million funding request for 2004 rejected by the Senate subcommittee that determines appropriations to the US Department of Commerce, of which NOAA is a part. The subcommittee proposes cutting off the SEC's funding completely.

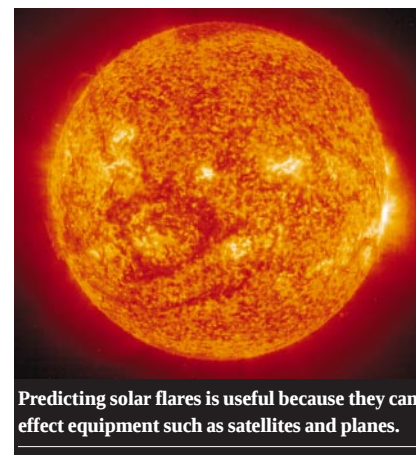
The House of Representatives' bill cuts the request by 40%, which would save the centre but cause "serious downsizing", according to the SEC's deputy director, Ron Zwickl.

The House and Senate bills will be reconciled within the next few weeks. The final appropriation is typically somewhere between the two figures, but in rare cases the full funding request can be restored.

The SEC is the primary US source of solar-activity forecasts, which are generated from satellite data and readings from ground-based solar telescopes and geomagnetic sensors. Because intense solar flares can harm equipment ranging from satellites in space to electrical power grids on the ground, the SEC's alerts go to a wide variety of customers, including NASA, the Pentagon, utility companies and airlines.

The congressional committees think that the public agencies should be paying for the data that they use. "If somebody really wants it badly enough in a tight budget environment, they need to step up to the plate," says Kevin Linskey of the Senate appropriations committee staff.

But the SEC's supporters, including Daniel Baker, director of the Laboratory for Atmospheric and Space Physics at the nearby University of Colorado, counter that NASA has traditionally been opposed to paying for monitoring networks. And the US Air Force, which uses the SEC's data to create its own space-weather forecasts,



Predicting solar flares is useful because they can effect equipment such as satellites and planes.

has not volunteered to fund the centre.

The American Astronomical Society has written letters of support and alerted its members to the SEC's plight. Advanced-technology company Lockheed Martin, United Airlines and several US utility firms have weighed in to defend the centre. Some NOAA officials hope that the centre could be saved by incorporating it into the agency's National Weather Service, or that the final bill will restore the funding. ■

Congress split over funding for 'safe' nuclear reactor

Geoff Brumfiel, Washington

Friends and foes of nuclear power in the United States are battling over a proposal to spend up to \$1.1 billion on an 'intrinsically safe' fission-reactor design.

Supporters of the plan, led by Senator Pete Domenici (Republican, New Mexico), have allocated \$30 million for it in the Senate version of next year's funding bill for the Department of Energy. But the House version makes no such provision, and the project's immediate fate will be determined at a conference between the two sides in the next few weeks.

The proposed reactor at the Idaho National Environmental and Engineering Laboratory near Idaho Falls would burn pellets of uranium fuel just 0.5 millimetres in diameter. Some five billion of the pellets would either be fused into billiard-ball-sized graphite spheres or encased in a honeycomb-like graphite structure. The system would be cooled by helium circulating at nearly 1,000 °C, which could directly drive turbogenerators. Advocates of the design claim that the helium could also be used to generate hydrogen for use as fuel.

"This new reactor will also be inherently safe," says Jim Lake, associate director of nuclear energy at the Idaho lab. In the event of coolant loss, the core will reach 15,000 °C but will not melt down, he says, because of graphite's high heat tolerance.

Critics counter that the design has other safety problems. "Graphite can't melt, but it can catch fire," says David Lochbaum, a nuclear engineer with the Union of Concerned Scientists, a non-profit environmental group.

Chris Paine of the Natural Resources Defense Council, another environmental group, doubts that generating large quantities of flammable hydrogen at a nuclear reactor will ever be safe or economic. "The capital costs of generating hydrogen from a nuclear reactor are absurdly high," he says. "It makes no sense."

Andy Kadak, a nuclear engineer at the Massachusetts Institute of Technology, says that a graphite fire is unlikely, and that hydrogen could be generated in buildings a safe distance from the reactor at off-peak times.

An energy bill that would recommend spending \$1.1 billion on the project over the next five years is being held up by disputes over unrelated issues, such as energy regulation. ■



Military physicians at the Dachau concentration camp tested the effects of icy water on prisoners.

Letters reveal scale of German agency's Nazi corruption

Quirin Schiermeier, Munich

Germany's scientific establishment was more thoroughly captivated by Nazi thinking than has ever been admitted, according to freshly collated documentary evidence.

The DFG, Germany's main research agency, is sponsoring a re-evaluation of its role in the period surrounding the Second World War. Science historians who met in Heidelberg last week to discuss this reported that the agency funded exploitative and unethical research between 1933 and 1945 across several disciplines, including medicine, physiology and psychiatry. Several participants said that any view that German science was not totally corrupted by Nazi ideology should be revised.

The agency has only recently begun to assess its Nazi past. It commissioned a book on the subject from the historian Notker Hammerstein, which was published in 1999 and roundly criticized for its apologetic tone and its tendency to downplay links between the DFG and medical war crimes (see *Nature* 398, 274; 1999).

The DFG then took a second, more comprehensive approach. It commissioned a working group to prepare for critical analysis tens of thousands of archived letters and reports relating to projects funded between 1920 and 1970. It then set up a five-year programme, comprising 14 peer-reviewed sub-projects, aimed at investigating DFG-funded research throughout the period.

"We have proven that the DFG was willingly involved in the full range of medical crimes during the Nazi era," says Wolfgang Eckart, head of the history of medicine at the University of Heidelberg and the organizer of last week's meeting.

Correspondence between grant applicants and Ferdinand Sauerbruch, then head

of the DFG's medical section, supports the view that the DFG leadership was fully informed about inhuman medical-research projects. In one letter, Otmar von Verschuer, director of the Kaiser Wilhelm Institute for Anthropology in Berlin, enthused about "excellent conditions for researchers in Auschwitz". Others express appreciation for the "wonderful brains" obtained from a euthanasia centre in Brandenburg-Görden, where at least 9,000 people, including many children, were killed.

The meeting was also given evidence that the DFG actively supported notorious human experiments. In the Dachau concentration camp, for example, tropical-disease expert Klaus Schilling infected 1,200 prisoners of war, including many Polish priests, in search of a malaria vaccine. About 400 died.

"During the war, the DFG deliberately subordinated medical ethics to the production of fast results," says Marion Hulverscheidt, a science historian at the University of Heidelberg. "Malaria experiments were supported against scientific advice, and even though valid animal models were available."

After Hitler came to power, the DFG became a de facto Nazi organization. Ordinary research was not eliminated, but it is a "myth" that science remained neutral and pluralistic, says Ulrich Herbert, a historian at the University of Freiburg, who co-chaired the investigation.

Allied intervention put a stop to medical crimes in May 1945, but Nazi ideas in the social sciences and humanities can be tracked until the 1960s, he says. "It becomes ever more evident that the DFG was a basic part of the system," says Guido Lammers, the DFG programme director who is overseeing the assessment. "There is a strong case for us to revise our history." ■

Channel hoppers land chemistry Nobel

Jim Giles

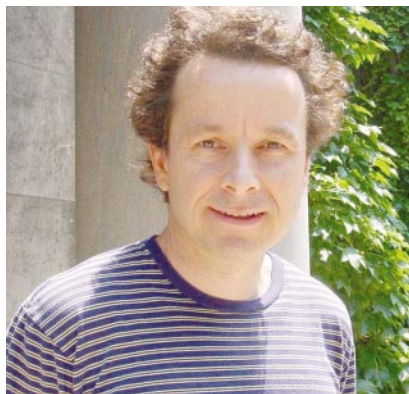
Two structural biologists credited with transforming our understanding of how cells work have been awarded the 2003 Nobel Prize in Chemistry.

Peter Agre of Johns Hopkins University in Baltimore and Roderick MacKinnon of the Rockefeller University in New York share the prize for experiments that revealed the intricate workings of the channels that allow ions and water to enter and leave cells.

MacKinnon is perhaps the most widely tipped Nobel winner of recent years. In a landmark 1998 paper (D. A. Doyle *et al. Science* **280**, 69–77; 1998), he and his colleagues provided the first detailed three-dimensional picture of a protein that acts as a channel to control the flow of potassium ions across cell membranes. This channel is immensely important to neuroscientists, as the flow of potassium ions helps to generate the voltage pulses that brain cells use to communicate.

Before MacKinnon's paper, many biologists had questioned whether the technique he used — X-ray crystallography — could be used to image membrane proteins. The method involves forming crystals of the protein to be studied, but researchers had previously failed to get large enough quantities of channel proteins for effective crystallization. MacKinnon experimented with potassium-channel proteins from many different organisms, eventually extracting enough from the bacterium *Streptomyces lividans*.

When he went on to describe how such channels open, close and prevent other ions from entering, many colleagues felt that he



Roderick MacKinnon (left) and Peter Agre's work on cell-membrane proteins has won them recognition.

deserved a Nobel. "His work is a *tour de force*," says John Walker, a structural biologist at the University of Cambridge, UK. "The award is absolutely spot on and richly deserved."

MacKinnon was on holiday when the prize was announced, and couldn't be informed directly by the Nobel prize committee. "My assistant called to tell me but I didn't believe him," says MacKinnon. "I had to do a Google search on 'Nobel prize'." When asked what he would do with the prize money, MacKinnon suggested that buying a mobile phone might be a good idea.

Agre's breakthrough came in 1988, when he cracked a problem that had survived more than a century of study — the way in which cells absorb and flush out water. Agre and his colleagues, while studying cell membranes, identified a previously unknown protein in red blood cells (B. M. Denker, B. L. Smith, F. P.



Kuhajda and P. Agre *J. Biol. Chem.* **263**, 15634–15642; 1988). "No one had seen it before, but we found that it was the fifth most abundant protein in the cell," says Agre. "That's like coming across a big town that's not on the map. It gets your attention."

The protein, now known as aquaporin 1, is a cell-membrane channel that lets water pass through, but not other substances. "We weren't looking for the water channel," says Agre. "We were very lucky."

"Agre's work changed the way we think about membranes," says Stefan Hohmann, a cell biologist at Gothenburg University in Sweden. The studies could have many medical spin-offs, he adds. Aquaporins play an important role in the brain swelling that can occur after stroke, for example, and blocking their function could help to reduce this effect. ■

♦ www.nobel.se

Economics prize for duo who resolved random results

Philip Ball

The Nobel prize for economics has been awarded to researchers whose mastery of mathematical methods could have consequences way beyond the 'dismal science'.

The prize was awarded to Robert Engle of New York University and Clive Granger of the University of California, San Diego, for their development of methods to analyse apparently random fluctuations in economic data.

The dynamics of economic statistics — including indices of national output, stock prices and exchange rates — pose a daunting challenge to scientific study. These indices tend to vary unpredictably over time, with periods of high volatility interspersed with calmer spells.

Engle, an American who trained as a physicist at Cornell University, devised statistical models to describe the plots of



Risk assessor: Robert Engle.

such indices over time — known as time series — using a concept that he called autoregressive conditional heteroskedasticity (ARCH). Models of market volatility based on the idea are now widely used by market analysts

and bankers to evaluate risk.

Welsh-born Granger, meanwhile, developed statistical methods that help analysts to deal with time series that fluctuate around a moving baseline, by combining them in a way that removes their long-term drift, through a phenomenon that he called co-integration. This makes such data amenable to conventional statistical



Clive Granger: on top of time series.

analysis, and helps to identify causal connections between data sets.

"If you are going to give a prize for economic time-series analysis, these are the two to give it to," says economist Paul Ormerod of London-based

company Volterra Consulting.

Besides economics, the winners' methods have found widespread application in the analysis of nonlinear systems in such diverse areas as climatology, physiology and physics. Granger's techniques have proved to be particularly versatile and have been used, for example, to assess human influences on climate change. ■

Gorilla count resumes as tensions ease in central African forests

London A census of the small band of mountain gorillas that inhabit the forests of the Virunga Massif in central Africa is being conducted for the first time in 14 years.

Such surveys used to take place every five years. But researchers have been prevented from visiting the animals' homeland — the forests that span Rwanda, the Democratic Republic of the Congo and Uganda — by conflict in the region.

The International Gorilla Conservation



In the black? A full survey is set to show whether central Africa's gorilla numbers are on the rise.

Program, a joint initiative involving the WWF conservation group and other environmental organizations, says that the last count, made in 1989, put the group's population at 320. Environmentalists fear that some gorillas may have died since then, as local people often kill them for their meat. But a partial census carried out in 2000 hinted that numbers may actually have increased.

Around 100 people are taking part in the count, which began last month. Organizers say that the survey's results will be available early next year.

Big Mac team serves up cheap supercomputer

Washington Computer scientists in Virginia have built one of the world's fastest supercomputers — out of 1,100 desktop machines and for just \$5 million, about one-twentieth of the price of a standard supercomputer.

Faculty members and students at Virginia Tech in Blacksburg linked together a mass of Apple G5 computers to make a machine, nicknamed Big Mac, that clocks in at 17.6 teraflops (17.6 trillion operations per second). Only five machines in the world can top speeds of ten teraflops — among this élite group is the Earth Simulator in

Yokohama, Japan, which has a peak rate of about 40 teraflops.

The Virginia team was hurrying to complete the project by 1 October, in order to qualify for a listing on the TOP500, a twice-yearly compilation of the world's fastest supercomputers. The listing, which is expected to be announced in mid-November, could help to attract top research projects to the facility.

NIH grant leads energy lab into protein project

Washington The Pacific Northwest National Laboratory in Washington state has won a \$10.2-million grant from the US National Institutes of Health (NIH) to set up a centre for basic research in proteomics.

The award is the largest NIH grant ever won by the lab, part of the Department of Energy, and marks a breakthrough for it. The lab has devoted much of its research to work connected to the adjacent Hanford site, which manufactured the United States' nuclear weapons during the cold war. But it has diversified into biology since opening an environmental molecular-sciences laboratory in 1996.

Half of the NIH grant will be used to develop high-precision mass-spectrometry and data-management tools that will help

researchers to describe the proteomes of mammalian cells quickly, according to Richard Smith, who will head the proteomics centre. The other half of the money will go towards funding collaborations with universities.

Telescope trains a MAGIC eye on the Universe

London The dark hearts of galaxies, supernovae, mysterious γ -ray bursts, and perhaps even a glimpse of the Big Bang itself, may soon be revealed by MAGIC — the Major Atmospheric Gamma Imaging Cherenkov telescope.

The telescope, inaugurated last week at the Roque de los Muchachos Observatory on La Palma, one of Spain's Canary Islands, has a 240-square-metre mirror that will be devoted to looking for γ -rays, the most energetic form of radiation in the Universe. MAGIC will focus on the weakest and therefore oldest of these γ -rays, and should see at least eight billion years into the past. It might even catch sight of γ -rays from the birth of the Universe 15 billion years ago.

Astronomers already have five ground-based γ -ray telescopes and one in orbit. But MAGIC will fill a gap by looking for particularly low-energy rays.

Fireball photo sparks explosion of debate

Washington Photographs of unidentified flying objects often cause a stir, but they don't usually end up on NASA's homepage. Last month, however, teenager Jonathan Burnett of Pencoed, Wales, snapped a mysterious fireball in the sky with his digital camera. He e-mailed the photo to NASA, who featured it as their 'astronomy picture of the day', describing it as "a sofa-sized rock that exploded as a daytime fireball".

If true, it would be one of the best-ever photos of a meteor trail. But experts are claiming that the snap actually shows sunlight reflecting from an unusual airplane contrail — perhaps that of the supersonic jet Concorde. The story is not yet over — others now claim to have seen the same event, and the image has generated a



flurry of e-mails among astronomers. NASA plans to publish more images of the phenomenon next week and to provide updates on the object's identity.

Scripps gets invitation to the sunshine state

San Diego Florida is offering the Scripps Research Institute, based in La Jolla, California, \$450 million to open a research laboratory near Palm Beach.

The funding, proposed on 8 October by state governor Jeb Bush, comprises \$140 million for a laboratory facility and \$310 million for operating costs for up to seven years.

Scripps opened its doors nearly 80 years ago as a medical clinic, and has evolved into a prominent research institute specializing in chemistry, immunology and cellular biology. It has 3,000 scientific personnel and an annual operating budget of \$280 million. Over the past 20 years, Scripps says that some 40 companies have spun off from it.

Bush has also called for a special session of the Florida Legislature on 20 October to consider several other economic projects for the state.



Time to choose

In some countries, transgenic plants are already a part of mainstream farming. Will the rest of the world soon follow suit?

Anyone attempting to predict the future is asking for trouble — especially at the cutting edge of science and technology, where the unexpected is the norm. But for the past few years, there has been one forecast that scientific soothsayers have been able to rely on: that controversy will rage over genetically modified (GM) crops.

So to say that agribiotech is now in for a volatile time might seem like an empty prediction. But there are good reasons to believe that the fight over GM crops is coming to a head.

Most European governments, realizing that their people have little enthusiasm for GM food, have been stalling on deciding whether to allow commercial plantings of transgenic crops. But following the introduction of a regulatory framework at the European Union (EU) level, they won't be able to stall for much longer. Decision time is dawning. If European countries say yes, they will face an onslaught from their own public. If they say no, the pro-GM US government will be spoiling for a fight. Already, the United

States has fired a warning shot across Europe's bow, lodging a complaint with the World Trade Organization over the EU's failure to open its markets to GM seeds and produce.

In Britain, the government has prepared for decision time by conducting by far the largest-ever trial of GM crops, seeking to gather as much evidence about their impact on biodiversity as possible. Those trials, along with extensive scientific evaluation and public consultation, are now coming to an end. But ultimately, the government's line may be influenced as much by Prime Minister Tony Blair's sagging popularity as by the scientific questions surrounding transgenic agriculture. In the following pages, *Nature* examines the pending British decision and places it in its wider international context (see page 656).

The outcome in Britain is bound to influence the debate in other countries where similar skirmishes are taking place. Around the world, environmentalists are battling with biotech-industry lobbyists to win over public opinion. In the developing world, the action

is poised to intensify, with sub-Saharan Africa emerging as an important new battleground. In other countries, such as China and India, the fight seems to be not so much 'GM or not GM', but rather between home-grown transgenic technologies and imports from agribiotech giants in the United States.

So far there is no clear winner. Our international survey of the extent of commercial cultivation of GM crops reveals a decidedly mixed picture (see page 658). Only a few countries have wholeheartedly embraced a transgenic future. But the agribiotech industry can point to several key markets where the prospects for GM farming are improving. Brazil, for instance, is the world's second-largest producer of soya beans — and there the tide seems to be turning in favour of GM varieties.

For now, our world map showing the market penetration of transgenic crops remains mostly blank. How quickly it fills up will depend on events and decisions that cannot be avoided for much longer.

Peter Aldhous, chief news and features editor.

P. DEAN/AGRIPICTURE.COM

Damned if they do, damned if they don't...

It's crunch time for agribiotech in Britain, as politicians rule on the planting of commercial transgenic crops. The world is watching, says Jim Giles.



Politicians don't often evoke pity. But it's hard not to feel some sympathy for the British ministers who must, over the coming months, decide on whether to give the green light for the commercial cultivation of genetically modified (GM) crops. They're in a 'no-win' situation.

On the one hand, the government wants to support the biotech industry, which it sees as a key component of Britain's future economic competitiveness. It also wants to appease its closest ally, the United States, which is already fuming over Europe's reluctance to embrace transgenic agriculture. But surveys have shown that, for now at least, the British public doesn't want to see GM crops grown commercially. And with the government's popularity dipping in the wake of the war in Iraq, brazenly defying public opinion is not an attractive option. "The British government is in a difficult position," concludes Simon Barber of EuropaBio, a Brussels-based body that represents Europe's biotech industry.

The arguments over transgenic agriculture have been played out more vociferously in Britain than in perhaps any other country. This ensures that the government's decision will resonate beyond the shores of the British Isles, colouring the debate over transgenic agriculture both at the European level and beyond.

Ministers must wish that they could kick the problem into touch and return to it when the political climate is more favourable. But matters are now coming to a head, thanks in part to a timetable of the government's own making. In 1999, as arguments over GM crops raged in

the media, the government announced that applications to begin commercial planting of herbicide-tolerant GM maize, oilseed rape (or canola) and beet would be put on hold, pending the results of 'farm-scale evaluations', taking several years, of the crops' impact on farmland biodiversity. Results from these huge experiments will be published this week.

Ask the people

But public concerns about GM agriculture run wider than the question of whether herbicide-tolerant crops will disrupt populations of weeds and invertebrates. There are anxieties about the safety of GM food, for example, and about whether transgenes will 'pollute' organic crops. So, over the past few months, the government has embarked upon an elaborate exercise in evaluation and public consultation. A scientific panel has reviewed the pros and cons of transgenic agriculture; Prime Minister Tony Blair's Strategy Unit has considered the economic case; and the 'GM Nation?' debate, involving more than 600 meetings, has sampled public opinion.

For the government, it has been a sobering exercise. The scientific panel, while pointing out areas of uncertainty that need further research, voiced no fundamental objection to transgenic agriculture. But if Blair hoped that the public would warm to the idea of GM farming, he was mistaken — the 'GM Nation?' debates revealed hostile attitudes towards the technology. Although the apparent depth of feeling will have been exaggerated by the presence of environmen-

tal activists at the meetings, other opinion polls have found little support for the commercialization of GM crops. And economic specialists have advised the government that there is little to gain by pressing ahead while consumers remain so suspicious.

Sources in the biotech industry accept that transgenic crops will not be a money-spinner in Britain until consumer opposition softens, but they are desperate for the government to send out a positive message by approving the commercialization, in principle, of herbicide-tolerant GM crops. Without such a move, they claim, investment and scientific talent will drift away to more favourable pastures.

Some senior plant-biotech researchers have already announced plans to leave Britain this year. "Public opposition has caused industry to bleed away, which reduces funding opportunities and options for the future employment of students," says Mark Tester of the University of Cambridge, who is shortly to join the Australian Centre for Plant Functional Genomics at the University of Adelaide. One company, Bayer CropScience of Hauxton, near Cambridge, suspended its field trials of GM crops late last month, complaining that its experimental plots could not be guaranteed protection from protesters intent on their destruction. Blocking or delaying the decision to approve transgenic crops for commercial use will exacerbate these trends. "It would send out a very negative signal," says Barber.

It would also widen the rift between the United States and Europe — something that Blair is anxious to avoid. The World Trade Organization is already considering a complaint brought by the United States against the European Union (EU) over its failure since 1998 to approve any new GM crop for





On trial: whatever the result of Britain's tests of transgenic oilseed rape (left and below left), the final decision will also be influenced by Tony Blair's desire to keep a sceptical public happy.



commercial planting or human consumption. Thanks to opposition from countries including Denmark, France, Greece, Italy and Luxembourg, some 20 pending applications have been left in limbo.

GM agriculture is not the only issue that splits the United States and Europe — the war in Iraq has also soured transatlantic relations. And as Blair calculates his political future, the two issues could become entwined. In Britain, doubts about the wisdom of invading Iraq are mounting, particularly in light of the death of David Kelly, a scientific expert at the Ministry of Defence. Kelly took his own life after being identified as the source of a BBC radio story alleging that the government had exaggerated intelligence on Iraq's weapons capability to win support for the war. The affair, now the subject of an independent inquiry, has dented public trust in the government. Against this

background, Blair will not want to be seen as disdainful of public opposition to GM crops.

The farm-scale evaluations could offer an escape route. If they suggest that GM herbicide-tolerant crops might damage the environment, biotech firms will have little basis for complaint if approval for commercial planting is denied or delayed. Some press reports have suggested that the farm-scale trials will indeed raise red flags against some of the tested crops. But with the scientists involved keeping the results under a strict embargo, the truth will only be unveiled this week.

Stalling tactics

If the farm-scale trials contain no show-stoppers, the best bet for the government would be to delay the decision until public disquiet over Iraq and Kelly has subsided. The government has already pledged to refer the results of the farm-scale evaluations to an expert scientific panel, the Advisory Committee on Releases to the Environment. Awaiting its advice will probably delay things until January. A second set of results from the trials, based on winter planting of oilseed rape, is due out in the middle of next year, and could be used to stall a decision yet further.

But by then, the government's hand may have been forced by events at the European level. Following the drafting of strict new rules on labelling of GM produce, together with requirements that should allow transgenic ingredients to be traced from farm to fork, the stalled process of approving GM crops for growth and sale in the EU is expected to resume around the turn of the year.

Even if developments in Brussels mean that the British government has little choice but to sanction the commercial planting of GM herbicide-tolerant crops in principle, it

might be able to placate public opinion in other ways. When the EU's provisions for tracing and labelling GM foods were agreed last July, member states were given the freedom to set their own 'coexistence' rules — designed to minimize cross-pollination between GM and non-GM crops. "Tough legislation in this area would knock commercial planting on its head," says a senior figure in one environmental organization.

The Agriculture and Environment Biotechnology Commission, another government advisory body, is currently grappling with the issue of coexistence, and is due to report any day soon. Opponents of transgenic farming argue, for example, that fields used to grow GM crops should subsequently be kept free of non-GM varieties for several years; the biotech industry retorts that such measures would make transgenic crops uneconomical.

New laws on liability could also make commercial planting financially unattractive. Environmental groups point out that forcing agribiotech companies to take legal responsibility if transgenes spread to organic crops, or have unforeseen effects on biodiversity, could provide another way for the government to approve the crops while ensuring that they are unlikely to be grown.

The biotech industry, meanwhile, may prefer a compromise that makes commercialization contingent on stringent scientific monitoring, while setting limits on the total area that can be cultivated. It is unclear whether such limits would be allowed under EU law, but they would suit agribiotech firms in the short term — especially as consumer demand is so low.

The government's eventual strategy may depend on Blair's popularity in the coming months, which in turn rests on the report of the inquiry into Kelly's death. "If Blair comes out badly, he won't want to take risks," says one British expert in food and environmental policy. Blair might then try to show that he is in tune with public opinion by bringing in rules that effectively shelve the introduction of transgenic agriculture. "This is a political decision, not a scientific one," agrees Tester.

If these predictions are correct, it would be an odd end to a debate in which environmental groups and the agribiotech industry have invested enormous effort. It would also be demoralizing for researchers who have devoted the past four years to the farm-scale evaluations — not to mention everyone involved in this summer's huge evaluation and consultation exercise. But that's politics for you. ■

Jim Giles is a reporter for *Nature* in London.

Farm-scale evaluations

► www.defra.gov.uk/environment/gm/fse/index.htm

Science review

► www.gmsciencedebate.org.uk

Economic review

► www.number-10.gov.uk/su/gm/index.htm

'GM Nation?' public debate

► www.gmnation.org.uk

Effective protection may alter the look of Venice

Those seeking to rescue Venice are caught between the devil and the deep-blue sea.

Sir — Your News Feature “Save our city!” (*Nature* **424**, 608–609; 2003) reviews the dilemma faced in selecting a large-scale engineering project to protect Venice, to be implemented and financed by the Italian government. I can see benefits in both of the two major projects discussed: the MOSE series of moveable gates, and the scheme to separate the city from its lagoon and the sea. Unfortunately, the first option is likely to serve only as a temporary protective measure for a century or two and cause collateral problems, such as increased pollution. The second, on the other hand, will permanently alter the configuration and setting of the Venice

we know, a city that is charmed and recognized by its lagoon.

Various ancient Greek coastal cities, some with canals and constructed centuries earlier than Venice, are now completely submerged, such as Herakleion off the coast of Egypt. They, too, faced problems of insufficient protection against water surges and sea-level rise, and subsidence caused by the building of monumental structures on inadequate foundations. Towards the end of their active history, when the sea level rose above the base of the buildings, structures toppled increasingly rapidly until the cities were finally submerged.

In geological terms, these cities were similar to Venice. Considering what befell them, the opposite of the geological tenet “the present is key to the past” may well apply to Venice. Perhaps a longer-term, more secure approach would be one involving comprehensive Netherlands-type ‘polder’ dyke constructions, with seawater being pumped out of the encircled city, water maintained in the canals, and those unique structures reinforced, where possible, by deep pilings.

Jean-Daniel Stanley

Geoarchaeology Program, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA

How do impact factors relate to the real world?

Sir — I have read with interest your debate on the impact of scientific work (for example, *Nature* **422**, 259–261; 2003, *Nature* **423**, 479–480 & 585; 2003 and *Nature* **424**, 14; 2003) but I do not agree with the position taken by Adam Łomnicki (*Nature* **424**, 487; 2003).

The problem is that Łomnicki and others established their careers at a time when competition among scientists had a different meaning. I am a young scientist and like everyone I would like to discover something interesting and new. However, when my colleagues and I discuss biological problems we always think about impact factors.

Recently, my friends wondered where to send their new paper — to journal X with an impact factor of 1.4, or to journal Y with an impact factor of 1.8. After 10 years, the average paper will be cited 14 times in journal X and 18 times in journal Y (assuming a constant citation rate, which is not the case, of course). Thus we compete furiously for just a few more citations, as the impact factor of most journals does not exceed three.

As Łomnicki states, citations are statistical processes, but even very good papers are cited only a few times. The question is whether the difference between 10 and 20 citations can really change our knowledge and understanding of nature.

Journals as well as scientists compete for impact factors. A journal that wants a higher impact factor has to encourage authors to publish in it, but with more papers coming in, more must be rejected. Generally, authors want to publish in

journals with the highest possible impact factor, but it is very difficult for most journals to improve their impact factors as most submitted papers have been rejected by better journals.

Impact factors provide an easy way to assess our achievements. But we do not know if small differences in citation number are valid indicators of our work, or how citations are related to the real world and solving its problems.

Piotr Skórka

Institute of Environmental Sciences, Jagiellonian University, ul. Gronostajowa 3, 30-387 Kraków, Poland

Pointless suffering of animals can be avoided

Sir — In the News article “Agony for researchers as mix-up forces retraction of ecstasy study” (*Nature* **425**, 109; 2003) the ‘agony’ of the embarrassed researchers was dwarfed by that of their primate subjects.

From an animal-welfare standpoint, the affair was tragic. From a scientific standpoint, it was pointless. After all, the researchers themselves have noted that extensive evidence from animal studies already shows that methylenedioxy-methamphetamine (MDMA, or ecstasy) is dangerous. If more proof of its recreational effects is needed, scientists should focus on the many humans who use such drugs.

Ethical and non-invasive studies are easily conducted and, in fact, are already being carried out to allow scientists to detect early signs of Parkinson’s disease or other neurological effects in drug users.

In research currently being conducted by J. H. Atkinson at the University of California, San Diego, human users of methamphetamine (speed) — the drug accidentally used in the ‘ecstasy study’ — are examined for neurological and cognitive function and other clinical features. A similar strategy would apply to the study of MDMA.

Neal D. Barnard

Physicians Committee for Responsible Medicine, 5100 Wisconsin Avenue, Suite 400, Washington DC 20016, USA

JET at risk if Europe can not afford to pay for ITER

Sir — I agree with much of what Richard Buttery says in his letter (*Nature* **424**, 995; 2003) about the Joint European Torus (JET), but I must point out that he has quoted my statement (*Nature* **424**, 4; 2003) out of context. I was asked by *Nature* to comment on the situation that might occur if there were no substantial increase in the fusion budget in the Seventh Euratom Framework Programme. I mentioned the possible closure of JET in the context of there being insufficient money to finance the European contribution to ITER.

A. M. Bradshaw

Max-Planck-Institut für Plasmaphysik, Boltzmannstrasse 2, Garching/Greifswald, Germany

correspondence

Contributions to Correspondence may be submitted to corres@nature.com. They should be no longer than 500 words, and ideally shorter. Published contributions are edited.

Science exiled

How the complexities of science suffer in the arena of public policy.

Politicizing Science: The Alchemy of Policymaking

edited by Michael Gough

Hoover Institution: 2003. 313 pp. \$15

Paul M. Grant

This is not suitable bedtime reading — not if you want to fall asleep, that is. Those who think that public policy should be based on sound science will be left in despair that such a goal can ever be achieved in the midst of the competing political interests endemic to modern industrialized democratic societies, exacerbated by scientific illiteracy on the part of both leadership and electorate.

Politicizing Science relates the personal trials and tribulations of 12 scientists whose careers were directly affected when their scientific advice conflicted with the political interests of those in power. Although several of its US stories pertain to the Clinton administration, the recent death of Edward Teller, bringing with it memories of the Oppenheimer affair, reminds us that conflicts between science and policy determination are ideologically invariant. These days, for instance, scientists who thoughtfully question the efficacy of the Bush administration's limited missile-defence initiative are not exactly welcome to spend the weekend on the president's ranch at Crawford, Texas. The essays in *Politicizing Science* illustrate that the risk of a given scientific issue becoming politicized depends on the difficulty of its proof and falsification, and on its perceived risks and potential benefits. A few examples from the book will illustrate the point.

The first essay is by William Happer, a professor of physics at Princeton University, who was director of basic energy sciences in the US Department of Energy during the administration of the first President Bush. Happer's tenure saw the 'discovery' of cold fusion, an event that rapidly became politicized. After all, who could ignore cashing in on the energy deliverance of mankind? Happer compares this episode with the Soviet agronomist Trofim Lysenko's subversion of genetic inheritance in favour of 'environmental determinism'. What they had in common was that each was clearly subject to Karl Popper's litmus test: scientists must attempt to falsify their hypotheses. Cold fusion was quickly disposed of in the West, where the litmus test could not easily be politically coloured. By contrast, the totalitarian Soviet Union protected the 'correct' interpretation of genetic inheritance until Stalin's demise.

When a hypothesis or assertions are precise and can be tested, a free society that demands full disclosure will eventually sort it all out. A recent example of just that was the



Power surge: science would have benefited if Richard Feynman had made it to the White House.

satisfactory resolution of last year's Bell Labs scandal (see *Nature* **419**, 419–421; 2002). But when the science gets 'fuzzy', as with carbon dioxide-forced global climate change, the effects of radiation or chemical agents, or bioengineering plants or animals for human purposes, opportunities for the politicization of science compound and abound.

Bernard Cohen, a nuclear physicist, has spent a large part of his later career on efforts, mostly unsuccessful, to attract coverage in the wide-circulation media of the facts about radiation and health. Especially revealing is his compilation of the numbers of stories relating to various 'everyday' accidents in *The New York Times* during the years 1974–78, before the 1979 crisis at the Three Mile Island nuclear plant. There were, on average, 120 reports per year on road accidents (US death toll: 50,000 per year), 50 on industrial accidents (12,000 killed each year in the United States) and 20 on asphyxiation (4,500 US deaths per year). For accidents involving radiation, there were 200 entries, despite the fact that none involved related illnesses or fatalities.

Robert Nilsson, a professor of toxicology, has worked for the Swedish Environmental Protection Agency, as well as that country's National Chemicals Inspectorate. Nilsson recounts the rise of politicized environmentalism in Sweden, enforced through a plethora of regulatory agencies created by a parliament long dominated by a single party and whose oversight seldom involves a single scientist. These agencies have a long reach, descending even to the composition of the

sand piles in playgrounds (crystalline silica has been identified as a low-risk carcinogen).

Roger Bate is concerned with the harm that the imposition of environmental standards devised for industrialized nations can do to developing societies. In particular, he focuses on how a ban on the use of the pesticide DDT in Africa has led to a disastrous re-emergence of malaria, which now kills 3,000 African children a day. DDT spraying in Africa began in the 1950s and greatly reduced the incidence of malaria. But environmental and economic pressures brought by developed nations led to its almost total discontinuance until recently, when attempts were begun to 'vector' its application to the walls of houses. However, the long-standing ban on DDT use means that almost none is now made, and there is a danger that the supply may run out.

But perhaps the most egregious example of political interference in the free and open discussion of unsettled scientific issues was the campaign conducted by an associate of the former senator and later vice-president Al Gore and members of his staff against Fred Singer and his colleagues, all vocal sceptics of a link between carbon dioxide emissions and climate change. Singer, a pioneer in the field of atmospheric measurements, was the first to predict that population growth would result in a greater concentration of methane, an important greenhouse gas. He is also a prolific writer on issues of the environment and climate change. In *Politicizing Science*, Singer recounts the pressure that was exerted on him to remove the name of the

distinguished oceanographer Roger Revelle from a paper they had jointly published expressing the view that the scientific basis for global warming was insufficient to justify immediate policy action at that time. A libel lawsuit filed by Singer against an individual working closely with Senator Gore was settled in 1994 by a full and complete admission by the individual of facts at issue, with a retraction and apology for the insinuations that led to the action being brought.

A common theme throughout the tales told in *Politicizing Science* is the notion of the precautionary principle or, more prosaically, "Look before you leap". We should always be conscious of the possible consequences of our scientific endeavours and cautious in the deployment of new applications of science. But, taken to its extreme, the precautionary principle can result, as Nilsson says of the situation in present-day Sweden, in "Look, but never leap".

The antidote to an overdose of the precautionary principle is the discipline of risk analysis and management, as argued many times by the contributors to this book. Risk analysis attempts to measure the risk of a given technology to the individual against its potential benefit to society as a whole. In the words of Chauncey Starr, one of the founders of probabilistic risk analysis, "the moral high ground assumed by well-meaning activists for single health causes may well be socially immoral when evaluated by the welfare of the total population".

Unfortunately, *Politicizing Science* lacks any discourse on how best to 'de-politicize' science. Here and there are hints that we need to 'get more involved'. Many, if not most, scientists are put off by the political process, forgetting that the pursuit of success in one's own profession is often quite political. We should seek and cultivate those rare individuals who combine the ability to carry out creative science with a personal populist appeal and an unshakeable belief in that paradigm of democracy: "You can't fool all of the people all of the time."

Perhaps the greatest native-born American scientist of my generation was Richard Feynman. His untimely death took from us not only a giant in physics but also a man of the people who was just beginning to capture the popular imagination and trust — and liked it. Think of a Feynman in the US Senate or White House, and the impatience he would have had with the cold-fusion imbroglio and the hand-wringing over the precautionary principle. Somewhere, sometime, another like Feynman is sure to surface. When that happens, let's campaign to get him or her elected to executive or legislative power in Washington, London, Moscow or Beijing, or wherever they're most needed. ■

Paul M. Grant is a science fellow at the Electric Power Research Institute, 3412 Hillview Avenue, Palo Alto, California 94303, USA.

Smashed into orbit

The Big Splat, or How Our Moon Came to Be

by Dana Mackenzie

John Wiley: 2003. 240 pp.

\$24.95 (US), \$38.95 (Canada), £17.50

Joseph A. Burns

Ours is a violent world and always has been. The whole Universe was produced in an unimaginable cataclysm, the Big Bang, about 13.7 billion years ago. Much later, 4.6 billion years ago, a nearby supernova explosion may have triggered the collapse of the protosolar cloud that became our Solar System.

This slim volume, written straightforwardly and engagingly by Dana Mackenzie, a mathematician turned freelance writer, describes how a collision produced Earth and the Moon. The giant-impact hypothesis — the 'Big Splat' of the title — maintains that an object larger than Mars slammed into proto-Earth during the final stages of its accumulation, giving birth to the Moon.

The Big Splat lays out ancient thoughts about the Moon's place in the cosmos, sketches the contributions of the greats of classical physics (Galileo, Kepler, Newton and Laplace), detours into topics such as celestial mechanics and navigation, recalls the Apollo programme, and finally describes the collisional model of the Moon's origin.

This historical tour turns out to be somewhat circular. The first known attempt to explain the Moon's origin occurred in the fifth century BC, when the Greek thinker Anaxagoras, after viewing a meteorite that had been observed falling from the sky, speculated that all celestial objects were glowing 'stone stars' flung off Earth. Apparently he got it right in the case of the Moon, but astronomy textbooks only a generation ago were

not so sure. They still listed three scenarios for lunar origin that had been developed in some mathematical detail a century earlier: the reclusive mathematician Edouard Roche contended that the two bodies were siblings, having 'co-accreted' as an orbiting binary; the scholarly George Darwin (son of Charles) promoted the idea that the Moon was our planet's child, having split off when a rotationally distorted primordial Earth became unstable; and later a cantankerous crackpot, T. J. J. See, argued that the Moon formed elsewhere, only to be snared intact by our planet.

These classical hypotheses were still debated vigorously as the space age dawned, even though the flaws of each were well recognized. Co-accretion would yield less angular momentum — the combined 'spin' of the Earth and Moon about one another — than the Earth–Moon system in fact has. Fission would require much more angular momentum than exists now, and there is no plausible explanation for how it could have started off. And the capture of an intact body is ridiculously improbable.

The Nobel Prize-winning chemist Harold Urey believed that the Moon, alone among the terrestrial bodies, was formed cold. To test this hypothesis of the Solar System's formation, Urey used his political influence in 1958 to get NASA's founding mission statement to focus on the origin of the Universe, which was "written plain to our eyes on the surface of the Moon". So Earth's only satellite became the primary scientific target of the US space programme. The lunar rocks returned by the Apollo and Luna missions showed that the Moon had much lower proportions of iron and volatiles than Earth, but their isotopic signatures had striking similarities to Earth's — as well as occasional differences. As is often the case, no model of origin survived its confrontation with data.

The final third of this book describes the



Splat! A giant object crashing into the proto-Earth may have given rise to the Moon.

W. K. HARTMANN

In retrospect

An Investigation of the Principles of Knowledge and of the Progress of Reason, from Sense to Science and Philosophy

by James Hutton

1794

Facsimile edition: Thoemmes: 1999.

Chosen by Paul N. Pearson

Following the publication of *On the Origin of Species* in 1859, Charles Darwin learned (and duly acknowledged) that two previous authors had anticipated the theory of evolution by natural selection. The first account to come to light was by Patrick Matthew, who had briefly outlined the mechanism in an appendix to his 1831 book *On Naval Timber and Arboriculture*. The second was by the physician William Wells, who had speculated on selection and human evolution in 1818.

But some 50 years ago, E. B. Bailey described a still older version of the selection theory from a 1797 manuscript by the geologist James Hutton — now chiefly famous for his early appreciation of geological time. Unfortunately, this work, entitled the *Elements of Agriculture*, never appeared in print. Now a more complete, published account has come to light from 1794.

An Investigation of the Principles of Knowledge is an intimidating philosophical treatise of three volumes, running to 2,138 pages in its original edition. Hutton's friend and biographer, John Playfair, presciently noted: "The great size of the book, and the obscurity which may justly be objected to many parts of it, have probably prevented it from being received as it deserves." The selection theory is the subject of an entire chapter in the second volume (see supplementary information). Hutton mused: "If an organised body is not in the situation and



The original *Origin*? James Hutton described natural selection as long ago as 1794.

circumstances best adapted to its sustenance and propagation, then, in conceiving an indefinite variety among the individuals of that species, we must be assured, that, on the one hand, those which depart most from the best adapted constitution, will be most liable to perish, while, on the other hand, those organised bodies, which most approach to the best constitution for the present circumstances, will be best adapted to continue, in preserving themselves and multiplying the individuals of their race."

For example, Hutton describes that in dogs that relied on "nothing but swiftness of foot and quickness of sight" for survival, "the most defective in respect of those necessary qualities, would be the most subject to perish, and that those who employed them in greatest perfection would be best preserved, consequently, would be those who would remain, to preserve themselves, and to continue the race". But if an acute sense of smell was "more necessary to the sustenance of the animal", then "the natural tendency of the race, acting upon the same principle of seminal variation, would be to

change the qualities of the animal, and to produce a race of well scented hounds, instead of those who catch their prey by swiftness". The same "principle of variation" must also influence "every species of plant, whether growing in a forest or a meadow".

Hutton was no mere armchair theorist. He came to his principle after experiments in plant and animal breeding, some of which are described in the *Elements of Agriculture* manuscript. These experiments led him to distinguish between seminal variation, which occurs in sexual reproduction and is heritable, and non-heritable variation, caused by the circumstances of soil and climate.

It is important to stress, however, that while he used the selection mechanism to explain the origin of varieties in nature, he specifically rejected the idea of evolution between species as a "romantic fantasy". Indeed, he was a deist and regarded the capacity of species to adapt to local conditions as an example of benevolent design in nature.

It may be more than coincidence that Wells, Matthew and Darwin were all educated in Hutton's home town of Edinburgh, a place famous for its scientific clubs and societies. Studies of Darwin's private notebooks have shown that he came to the selection principle independently of earlier authors, as he always maintained. But it seems possible that a half-forgotten concept from his student days resurfaced afresh in his mind as he struggled to explain the observations of species and varieties compiled on the voyage of the *Beagle*.

Paul N. Pearson is at the School of Earth, Ocean and Planetary Sciences, Cardiff University, Cardiff CF10 3YE, UK.

Supplementary information The full text of Hutton's chapter from *The Principles of Knowledge* and other relevant extracts from *Elements of Agriculture* are available on *Nature's* website.

development of the now-preferred scenario for how the Moon came to be. By the mid-1970s the analytical ideas of the Soviet cosmogonist V. S. Safronov about the role of impacts in forming the planets of our Solar System had travelled west, where they were tested and extended numerically by George Wetherill and by the Planetary Science Institute. Bill Hartmann and Don Davis, from the latter group, realized that the final objects to accumulate into the terrestrial planets must have been massive and would have careened through the inner Solar System. They concluded that the Moon could have been born from the final collision of such an object, and argued that the material thrown off proto-Earth would have been from its iron-poor surface layers and that volatiles would have boiled off, explaining the Moon's gross composition.

Independently, Al Cameron and Bill Ward noted that the angular momentum of the entire Earth–Moon system required an impact from at least a Mars-sized projectile, and began to analyse the likely evolution of the flattened, Earth-circling cloud of

vaporized material that any giant collision would have generated. The Moon, accumulated from this orbiting debris through much the same processes as the planets themselves, would contain material from both Earth's mantle and the projectile. Despite their plausibility and close match with known facts, these works attracted little attention and languished along with all lunar studies during the early 1980s.

Languished, that is, until a conference in 1984, held in Kona, Hawaii, brought together dynamicists, cosmochemists and geophysicists to consider the Moon's origin. By this time, impacts were accepted as shapers of life, following the proposal of Luis and Walter Alvarez that linked an asteroid collision with the dinosaurs' demise and other mass extinctions at the end of the Cretaceous period. As the conference proceeded, it became clear that a consensus had silently emerged in the various disciplines: each, unaware of the other, favoured a collisional beginning for the Moon.

In the past two decades, there have been

increasingly sophisticated simulations of a massive impact and the evolution of the resulting debris disk, and lunar samples and meteorites have been scrutinized. The giant-impact hypothesis does well in explaining the observed physical, thermal and geochemical properties of the Earth–Moon system. Plausible modifications to the original theory may overcome the remaining faults: contemporary models suggest that Earth was not fully grown when it was struck by an even larger projectile than was previously thought.

Besides telling an interesting tale well and elucidating how science progresses, Mackenzie's book emphasizes the fact that impacts have been the primary creative and destructive process throughout the history of the Solar System. Some today wonder whether the final violence that ends civilization will be another such collision — rather than a catastrophe of our own making. ■

Joseph A. Burns is in the Departments of Theoretical and Applied Mechanics and of Astronomy, Cornell University, Ithaca, New York 14853, USA.

Avoiding the shrink

Ray H. Baughman

When compressed in a vice, the sides of a material are expected to bulge. Conversely, stretching is expected to make a material thinner — as for the rubber band on your desk. Most materials exhibit these dimensional changes in lateral directions to decrease the volume change produced by the applied strain. But materials do exist that expand in at least one lateral direction when stretched. Known as auxetic materials, they have recently gone from being rarely recognized to appearing rather commonplace. In fact, much of the crystalline matter in the Universe may be in auxetic phases, from most cubic phases of metals to super-dense crystals thought to comprise the outer crust of neutron stars and the cores of white dwarfs. These auxetic materials exist across an enormous range of conditions — from near absolute zero and densities of approximately $10^{-15} \text{ g cm}^{-3}$ for plasma ion crystals to above 10^6 K and densities of $10^{11} \text{ g cm}^{-3}$ for proposed crystals in stars.

The Poisson ratio (defined as -1 times the ratio of lateral to applied elastic strains) describes the behaviour of stretched materials — positive Poisson ratios characterize the expected normal behaviour, whereas auxetic materials have a negative value. Although Poisson's ratio for cubic phases must be between -1 and $+2$, there is no theoretical limitation on this ratio for materials with less internal symmetry. Although few crystals have negative Poisson's ratios for all stretch and lateral directions, such behaviour can be obtained for foams of most materials.

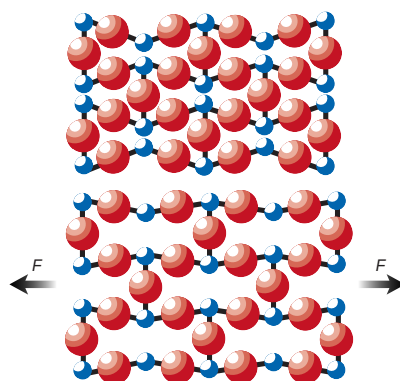
The behaviour of some auxetic materials seems more fitting for Alice's Wonderland than for the real world. After an initial elongation, stretching the porous polytetrafluoroethylene used as artificial arteries generates up to an 11-fold higher relative expansion in one lateral direction. Although the direct effect of a negative Poisson's ratio is to cause an expansion that decreases density during stretching, a small fraction of auxetic crystals actually increase in density when stretched. This behaviour is quite rare, but all 13 observed stretch-densified crystals (out of 500 investigated crystals) are auxetic.

The answer to this paradox is that very large positive Poisson's ratios are made thermodynamically possible by the existence of negative Poisson's ratios. Indeed, cubic metals that have the most positive Poisson's ratio for one lateral direction also have the most negative Poisson's ratio for the second lateral direction. One origin of this association between negative and positive Poisson's ratios is illustrated in the picture. Applying

a force F elongates the sheet by causing bond angle changes at the blue atoms, which produces a large lateral expansion (corresponding to a negative Poisson's ratio) and opening of void space in the sheet. Neighbouring sheets of atoms, above and below a central sheet, can then move closer to partially occupy this void space — thereby increasing the density of the material (corresponding to a positive Poisson's ratio).

Auxetic materials could be used as strain amplifiers if the strain-amplification factor (-1 times the Poisson's ratio) is larger in magnitude than one. Strain amplification close to the maximum Poisson's ratio of $+2$ occurs for cubic phases, and even this small amplification could be useful. However, auxetic materials with lower symmetry can have giant strain amplification factors, for example porous polytetrafluoroethylene has a Poisson's ratio of up to -12 , which is about 40 times larger than that for most materials. Amplification of the effect of hydrostatic pressure on a linear dimension is possible for the small fraction of auxetic crystals that are stretch densified, as these crystals also must have a negative linear compressibility — and a linear compressibility can exceed the volumetric compressibility if a linear compressibility is negative. In fact, a linear compressibility of auxetic lanthanum niobate crystals in one direction is 9.6 times greater than its average linear compressibility. Likewise, stretching (or compressing) a known auxetic polymer produces over a ten-fold larger magnitude volumetric strain.

There are various possible applications of these amplification effects. A slab of auxetic material could replace cumbersome lever systems currently used for amplification, which degrade frequency response, thereby amplifying an actuator-generated strain by over an order of magnitude. As the structural elements providing auxetic behaviour can have molecular dimensions, auxetic materials



Stretching some rare types of auxetic materials surprisingly makes them denser.

Auxetic materials

Materials that expand laterally when stretched can act as molecular-scale strain amplifiers. This amplification might be exploited in nature and in future technologies.

could be used in molecular-scale amplifiers. Large negative linear compressibilities — the required inverse effect to stretch densification — might find application for interferometric optical pressure sensors, where pressure-induced increases in both a dimension and density could avoid the usual partial cancellation of phase shifts due to changes in the refractive index and physical path length. Optical lenses piezoelectrically driven to provide very large changes in refractive index also seem achievable, as refractive index is sensitive to volume and a tensile strain can produce a ten times larger volumetric strain.

Such mechanical amplification associated with auxetic materials might explain why cell growth and metabolism are affected by minute pressures of approximately 10 kPa, such as those experienced when biological organisms are submerged in a metre of water. For example, lipid bilayers have been observed to increase in thickness when hydrostatically compressed (signifying a negative linear compressibility and stretch densification), and a negative Poisson's ratio has been calculated for membranes found in the cytoskeleton of red blood cells. As phase stability requires that the sum of all linear compressibilities for any material is positive, a material with a negative linear compressibility must be incompressible in certain directions (between directions of positive and negative linear compressibilities). Use of this feature might help marine animals preserve dimensions key for biological functions while travelling to the depths of the ocean. ■

Ray H. Baughman is in the NanoTech Institute and Department of Chemistry, University of Texas, Dallas, Richardson, Texas 75083-0688, USA.

FURTHER READING

- Baughman, R. H. *et al. Science* **288**, 2018–2022 (2000).
- Baughman, R. H., Stafström, S., Cui, C. & Dantas, S. O. *Science* **279**, 1522–1524 (1998).
- Bowick, M., Cacciuto, A., Thorleifsson, G. & Travesset, A. *Phys. Rev. Lett.* **87**, 148103–148106 (2001).
- Evans, K. E. *Adv. Mater.* **12**, 617–628 (2000).
- Gibson, L. J. & Ashby, M. *Cellular Solids* (Pergamon, Oxford, 1988).
- Lakes, R. *Adv. Mater.* **5**, 293–296 (1993).
- Macdonald, A. G. & Fraser, P. J. *Comp. Biochem. Physiol. A* **122**, 13–36 (1999).

The coelacanth of frogs

S. Blair Hedges

A frog that lives in the mountains of southern India is a rare breed indeed: it is a new species that merits the establishment of a new family. Moreover, this is a discovery with considerable biogeographical significance.

The discovery of a living coelacanth in 1938 captured public attention because it represented an ancient lineage of fishes thought to have been extinct for some 80 million years¹. Now, a living amphibian with unusually deep evolutionary roots has been discovered in India. Writing on page 711 of this issue², Biju and Bossuyt describe this odd-looking species of frog, which was collected in the Western Ghats Mountains of southern India. The characteristics that seem strange to a non-herpetologist — a bloated body, stubby limbs, tiny eyes and protruding snout (Fig. 1) — are not uncommon in burrowing frogs. However, its internal anatomy and DNA sequence data show that this species represents a deep branch in the family tree of frogs. Its closest relatives live in the Seychelles, 3,000 kilometres south of India, near Madagascar. Appropriately, the authors place their new species in a new family.

Just how significant is the discovery of another family of frogs? Only 29 families are known, encompassing the approximately 4,800 known species³. Most of these families were named by the mid-1800s, and the last discovery of a species of frog belonging to a new family, as opposed to merely a taxonomic rearrangement, was in 1926 (ref. 3). All others date to the 1700s and 1800s, making this a once-in-a-century find. Moreover, according to fossils and evolutionary 'clocks' devised using molecular data, families of frogs are about as ancient as orders and superorders of mammals, having diverged

from one another during the heyday of the dinosaurs in the Mesozoic era (251–65 million years ago)^{2,4}.

This discovery also draws attention to our incomplete knowledge of biological diversity, even at the higher taxonomic levels. The home of this plump amphibian, the Western Ghats, is one of the eight 'hottest hot spots' of biodiversity in the world⁵, meaning that many species occur there that are found nowhere else. Like most other hot spots, it was a region once covered with tropical forests. But pressures from human activities, such as farming, have reduced the forests to less than 10% of their former extent⁵.

Biologists are racing to survey and discover species in hot spots before they disappear. Unfortunately, fieldwork can be dangerous (diseases, guerrilla wars, venomous animals), and greater efforts are now required to reach unaltered habitats, such as the tops of mountains. Moreover, some governments are afraid of losing their countries' genetic resources, and have been discouraging foreign scientists from collecting plants and animals. To complete a gloomy picture, taxonomy in general has become an unpopular career choice. Nonetheless, extraordinary discoveries such as this frog show that there is an urgent need for more biotic surveys.

Of great interest to biogeographers is the finding that the new species is most closely related to Seychellean frogs of the family

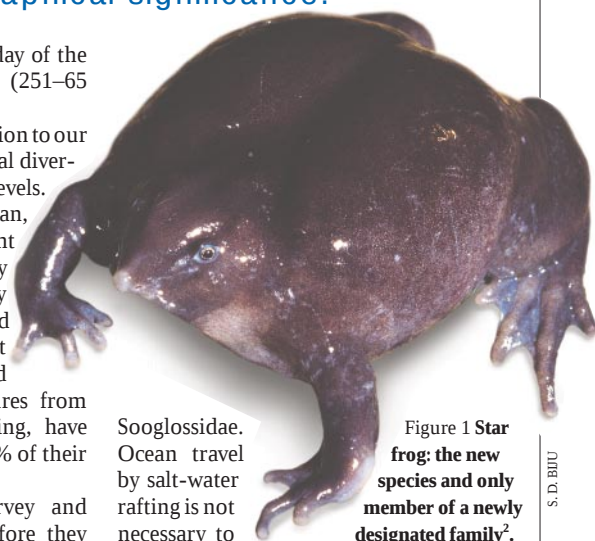


Figure 1 Star frog: the new species and only member of a newly designated family².

Sooglossidae. Ocean travel by salt-water rafting is not necessary to explain this relationship, because models of continental drift in the Mesozoic place India and the southern land areas together. The southern supercontinent Gondwana began breaking up 160 million years ago, separating a western continent of South America and Africa from an eastern continent of India, Madagascar, the Seychelles, Antarctica and Australia. Thereafter, the eastern continent continued to break apart, losing Antarctica–Australia (130 million years ago), Madagascar (90 million years ago), and finally the Seychelles (65 million years ago) as India continued northwards. India's collision with Asia, 55 million years ago, created the Himalayas^{6–8}. If this geological history is correct, India should have been a 'biotic ferry' with a passenger list of distinctive plant and animal groups that evolved in isolation for tens of millions of years (Fig. 2a).

Oddly, however, the late Mesozoic fossil record of India does not support a biotic ferry model. Instead, it reveals organisms with close relatives in Africa, South America and Asia^{7,8}, including dinosaurs, lizards, frogs and mammals. Because the geological data allow some flexibility in reconstructing palaeogeography, new models have been proposed that incorporate late Mesozoic land bridges between India and other areas, especially Africa (Fig. 2b,c)^{7,8}.

How does the new species of frog from India bear on these biogeographical models? If it diverged from the Seychellean frogs (sooglossids) as early as molecular clocks indicate, 130 million years ago², continental

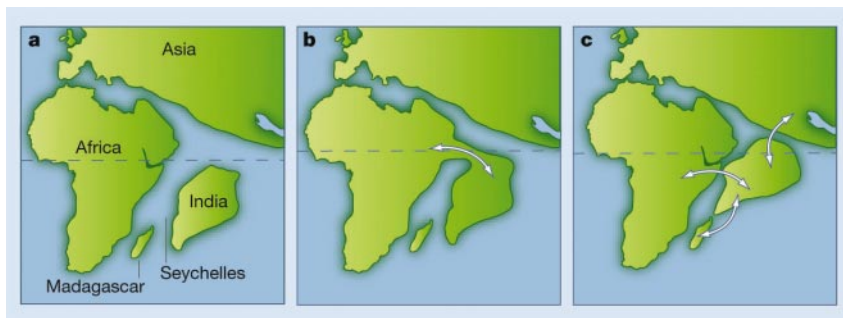


Figure 2 Possible Indian odysseys: different models for the position of India approximately 65 million years ago. a, The standard 'biotic ferry' model showing India isolated by large expanses of water⁶. b, A limited 'biotic (land) bridge' model incorporating a narrow connection (Greater Somalia) with Africa⁷. c, Another biotic bridge model assuming a different longitudinal position for India and showing connections with Madagascar, Africa and Asia⁸. The discovery of a new species of frog in India lends support to a biotic ferry model, but the fossil evidence of other animals favours the existence of land bridges.

breakup would not directly explain its origin — India did not split from the Seychelles until 65 million years later. However, the apparent isolation of Biju and Bossuyt's frog family in India supports the biotic ferry model. Molecular-clock studies of other living groups of plants⁹ and animals^{10,11}, including caecilians (limbless amphibians), also indicate that India developed a unique biota during its northward trek.

But why does the current biota reflect such isolation while the late Mesozoic fossils of India indicate past land connections ('biotic bridges')? Perhaps those bridges were more like chains of islands that allowed some — but not all — groups to disperse, as occurred in the past history of plant and animal interchange between North and South America¹².

The discovery of this remarkable new species adds to growing evidence of past isolation in the biogeographical history of India. Nonetheless, it is unclear why India's Mesozoic partner Madagascar lacks some major groups of vertebrates, such as caecilians and representatives of the new frog family, when evolutionary analyses indicate

that they should have been there in the past. Clearly, there is a need for more fossil collections and investigation of living faunas, and for refined molecular clocks, to better understand how continental drift influenced India's biota.

S. Blair Hedges is in the NASA Astrobiology Institute and Department of Biology, Pennsylvania State University, University Park, Pennsylvania 16802, USA.

e-mail: sbh1@psu.edu

1. Thomson, K. S. *Living Fossil: The Story of the Coelacanth* (Norton, New York, 1992).
2. Biju, S. D. & Bossuyt, F. *Nature* **425**, 711–714 (2003).
3. Frost, D. R. *Amphibian Species of the World: An Online Reference Version 2.21* (15 July 2002). Available at <http://research.amnh.org/herpetology/amphibia> (2003).
4. Hedges, S. B. *Nature Rev. Genet.* **3**, 838–849 (2002).
5. Myers, N., Mittermeier, R. A., Mittermeier, C. G., Da Fonseca, G. A. B. & Kent, J. *Nature* **403**, 853–858 (2000).
6. Scotese, C. R. *Atlas of Earth History Vol. 1, Paleogeography* (Paleomap Project, Arlington, Texas, 2001).
7. Chatterjee, S. & Scotese, C. R. *Proc. Indian Natl Sci. Acad.* **65A**, 397–425 (1999).
8. Briggs, J. C. *J. Biogeogr.* **30**, 381–388 (2003).
9. Conti, E., Eriksson, T., Schönenberger, J., Sytsma, K. J. & Baum, D. A. *Evolution* **56**, 1931–1942 (2002).
10. Bossuyt, F. & Milinkovitch, M. C. *Science* **292**, 93–95 (2002).
11. Gower, D. J. et al. *Proc. R. Soc. Lond. B* **269**, 1563–1569 (2002).
12. Simpson, G. G. *Splendid Isolation* (Yale Univ. Press, New Haven, 1980).

Earthquakes

Good tidings

Christopher H. Scholz

Tidal stresses in the Earth's crust don't seem to influence earthquakes. Water wells, on the other hand, seem strangely sensitive to seismic activity. Explanations are now proposed.

The respective absence and presence of two phenomena associated with earthquakes has been puzzling geophysicists for more than a century. One is the general lack of correlation between earthquakes produced by tectonic forces and the 'solid Earth tides', which are caused by the oscillating stresses created in Earth's crust by the gravitational forces exerted by the Sun and the Moon. The other is that some water wells are extraordinarily sensitive to the seismic waves of distant earthquakes. Papers by Beeler and Lockner¹ and by Brodsky et al.², both in the *Journal of Geophysical Research*, help to explain these phenomena — and perhaps a few others as well.

If earthquakes have simple behaviour, in which stress on a fault builds up to some threshold at which the fault fails, then one would expect their occurrence to correlate with the daily Earth tides. After many increasingly sophisticated studies, no such general correlation has been found. Yet it has been recently recognized that changes in 'static stress' from earthquakes can trigger other earthquakes³, even when the stress change is as low as 1 kilopascal (ref. 4) — which is roughly the same mag-

nitude of effect associated with Earth tides.

Beeler and Lockner¹ conducted rock friction experiments in the laboratory to simulate the situation of a small sinusoidal loading (the Earth tide) being superimposed on linear loading (a fault being loaded tectonically). The laboratory equivalent of the earthquake cycle is 'stick-slip' events, in which frictional stress builds up at a 'fault' until its adjacent sides begin to slip and a slip instability occurs (Fig. 1a). Beeler and Lockner mapped the amplitude of the oscillating stress necessary to produce correlation with stick-slip events as a function of the oscillation period. They found two regimes. For oscillations with periods greater than a critical time, the correlation amplitude decreases as $1/f$, where f is the frequency of the oscillation. This is just as would be expected from a model — the Coulomb threshold model — which assumes that failure occurs at the peak stress. At periods shorter than the critical time, they found that the correlation amplitudes become frequency independent and are orders of magnitude larger than expected from the Coulomb model (Fig. 1b).

The key point is that the stick-slip instability in rock friction is not abrupt. Rather, it

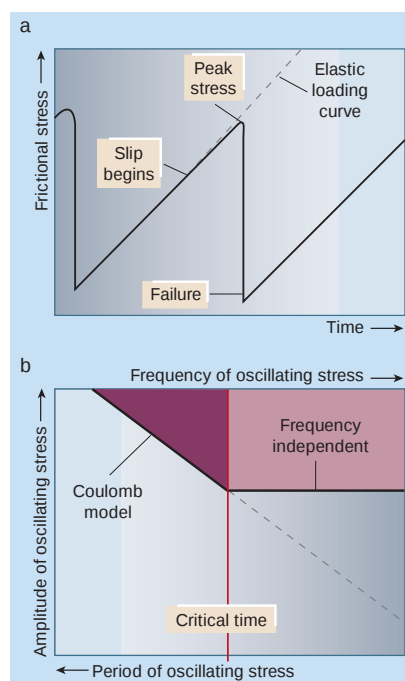


Figure 1 Stick-slip and Earth tides. a, The stick-slip process describes the changing level of frictional stress at a fault over time. Adjacent sides of the fault stick together as frictional stress builds up, but then begin to slip as the rate of stress change starts to diverge from the elastic loading curve. The peak stress value is reached slightly ahead of the point of failure. b, Beeler and Lockner¹ looked at the amplitude of oscillating stress, mimicking Earth tides, that reproduces stick-slip events. For oscillations with periods greater than a critical time, the correlation amplitude follows the Coulomb threshold, decreasing in inverse proportion to the oscillation frequency (all scales are logarithmic in this plot). But for short-period, high-frequency oscillations, the threshold for correlation no longer changes with frequency, and is considerably higher than would be expected in the Coulomb model.

follows a nucleation phase, in which the rate of slip increases as an inverse power law of the time to failure, leading to a peak in stress. Fault failure occurs following the peak. The critical time dividing the two correlation regimes corresponds to the duration of nucleation, which is inversely proportional to the linear stressing rate. This nucleation requires that frictional resistance increases with slip velocity and decreases with slip displacement. Beeler and Lockner have constructed a friction law incorporating these properties — a simple form of rate-state variable friction⁵ — and show that it correctly predicts the behaviour in the high-frequency regime.

At periods greater than the nucleation time, the nucleation has no effect. But at shorter periods it greatly dampens the triggering effect, so that much higher amplitudes

are required to produce a correlation. For instance, given the stressing rate on the San Andreas fault, the nucleation time exceeds a year. So a static-stress change of the same amplitude as Earth tides can trigger earthquakes. But the tides are well into the high-frequency regime and are far too low in amplitude to do so.

In their work, Brodsky *et al.*² studied the response to earthquakes of a well in Oregon. The well is set in granite, and amplifies the long-period seismic waves produced in earthquakes by a factor of about 300: a long, thin crack intersects the borehole at depth and acts as a bellows, drawing water in and out of the well when compressed and expanded by seismic waves. Curiously, however, the surface waves of large earthquakes as much as 3,000 km distant produce sustained changes in water level of about 10 cm.

Brodsky and her colleagues put forward a model in which groundwater flow through the crack deposits weathering material that clogs the crack. They propose that the passage of surface waves dislodges the blockage, allowing the pressure in the crack to re-equilibrate, resulting in a step in water level in the well. Two independent corroborations make this a convincing argument. Brodsky *et al.* observed that the amplification factor of the well increased by several orders of magnitude during a step, suggesting that the length of crack in communication with the well had greatly increased during the step. Furthermore, although the 1999 Oaxaca earthquake in Mexico (magnitude 7.4) produced an 11-cm step, the Hector Mine earthquake in California (magnitude 7.1), which followed it 15 days later, did not. The implication is that insufficient time had passed to form a new blockage.

The same physics embodied in Beeler and Lockner's analysis¹ with Earth tides explains why dynamic triggering by seismic waves requires much greater stress amplitudes (1–10 megapascals) than does static triggering⁶. As a result, static triggering dominates and dynamic triggering can be distinguished only under special circumstances⁷. The exception is volcanic areas, where earthquakes occur through the stress effects produced by magma activity on fractures in the rock. In this case, surface waves, similar in amplitude to those that produced steps in the Oregon well, can trigger earthquakes at large distances⁸. Brodsky *et al.* suggest that the unclogging mechanism may be responsible for that phenomenon also, because such steps would result in static-stress changes.

Earthquakes in volcanic regions differ from tectonic earthquakes in another way: earthquake 'swarms' in such areas have often been found to correlate with Earth tides⁹. This could happen because the stressing rate during swarms is much greater than tectonic

rates, so that the nucleation time becomes less than the tidal period. This is implied by an analysis of dynamic triggering by seismic waves in a volcanic region, which shows a threshold that depends on $1/f$, the characteristic of the low-frequency regime¹⁰. Otherwise, perhaps some crack–fluid interaction, such as that found by Brodsky *et al.*², could be triggered by the tides.

Christopher H. Scholz is at the Lamont–Doherty Earth Observatory, Columbia University, Palisades, New York 10964-8000, USA.
e-mail: scholz@ldeo.columbia.edu

Proteomics

Where's Waldo in yeast?

James A. Wohlschlegel and John R. Yates

Research in yeast provides the tools and benchmarks for a wide sweep of biology. The latest results reveal the most complete picture yet of the levels and locations of protein production in the organism.

Biologists have just completed a round of 'Where's Waldo?' with proteins of the yeast *Saccharomyces cerevisiae*. 'Where's Waldo?' is a popular children's game that entails finding every figure of Waldo in an intricately drawn picture. In some areas of the picture, Waldo blends in well with the background and is difficult to find; in others he stands out. Two papers in this issue (pages 686 and 737)^{1,2} describe large-scale studies of the yeast proteome, which had the goals of quantifying the levels of every protein expressed in yeast cells and determining the subcellular compartment in which the protein resides.

Saccharomyces cerevisiae was the first eukaryote — the type of organism characterized by a nucleus and membrane-bound organelles, which also includes humans — to have its genome sequenced³. Work with this organism has since led the way in functional genomics. Experiments pioneered in yeast have set the standard for the global analysis of cellular processes and paved the way for similar approaches in other organisms. They have also generated genome-wide collections of reagents that have been tremendously valuable.

Open reading frames (ORFs) are commonly the centre of attention in genome biology. These are stretches of DNA that have the characteristics of protein-coding capacity; that is, they may be genes. Collections of yeast strains now exist in which the expected ORFs have been either deleted or fused to various protein tags^{4,5}. Arrays have been created by using yeast strains expressing proteins that carry so-called affinity tags, allowing large numbers of proteins to be rapidly purified, then immobilized on a solid support⁶. Large-scale studies involving various techniques — protein arrays, and yeast

1. Beeler, N. M. & Lockner, D. A. *J. Geophys. Res.* **B 108**, doi:10.1029/2001JB001518 (2003).
2. Brodsky, E. E., Roeloffs, E., Woodcock, D., Gall, I. & Manga, M. *J. Geophys. Res.* **B 108**, doi:10.1029/2002JB002321 (2003).
3. Stein, R. S. *Nature* **402**, 605–609 (1999).
4. Ziv, A. & Rubin, A. M. *J. Geophys. Res.* **105**, 13631–13642 (2000).
5. Scholz, C. H. *Nature* **391**, 37–42 (1998).
6. Perfettini, H., Schmittbuhl, J. & Cochard, A. *J. Geophys. Res.* **B 108**, doi:10.1029/2002JB001805 (2003).
7. Kilb, D., Gombert, J. & Bodin, J. *Nature* **408**, 570–574 (2000).
8. Hill, D. P. *et al. Science* **260**, 1617–1623 (1993).
9. Kasahara, J. *Science* **297**, 348–349 (2002).
10. Gombert, J. & Davis, S. J. *J. Geophys. Res.* **101**, 733–750 (1996).

two-hybrid or co-immunoprecipitation assays — have revealed the identities of proteins that interact with individual proteins, large macromolecular complexes, or even specific small molecules^{7–10}.

All in all, yeast biologists have led the charge in developing approaches to understanding eukaryotic genomes. Huh *et al.*¹ and Ghaemmaghami *et al.*² continue that tradition. Their goal was to tag and study the gene products of all recognized ORFs in the yeast genome. A key component of these studies was the tagging method used: artificially altering a protein's expression level can lead to results, such as mislocalization, that do not reflect its characteristics when it is expressed normally.

In technical terms, Huh *et al.* and Ghaemmaghami *et al.* used homologous recombination to integrate a DNA sequence, encoding either a tandem affinity purification tag (TAP) or green fluorescent protein (GFP), in-frame with the 3'-end of the coding sequence of each gene in its original chromosomal location. Because a gene's promoter and upstream regulatory sequences are not affected in this approach, it is likely that the behaviour of these fusion genes is nearly identical to that of their normal counterparts.

Ghaemmaghami *et al.*² created a collection of yeast strains in which each annotated ORF was fused to a TAP tag. In turn, the tag was used to measure the absolute expression levels of each protein by a technique known as quantitative immunoblotting. They found that around 80% of the ORFs are expressed in growing yeast cells at levels ranging from 50 to 1,000,000 copies per cell. Knowing what portion of the yeast genome is expressed, and the absolute protein abundance, will help in verifying the accuracy and detection limits

theory. These include the velocity of the emitting material in the direction of the line of sight to the Sun; the magnetic-field strength; and the inclination of the local magnetic-field vector (normal to the solar surface and in the azimuthal plane at the level of both the photosphere and upper chromosphere).

Solanki *et al.*¹ have devised a technique for reconstructing the magnetic-field geometry of the observed regions from these data. Their results confirm much of the earlier work that was focused on the structure of the magnetic field on the solar surface, but also shed new light on how the magnetic field changes with height above the solar surface. At the level of the photosphere, they find that the field is highly irregular, which is consistent with earlier measurements using other techniques (including the aforementioned Zeeman effect). At the level of the chromosphere, the field seems to be much more regular.

What is perhaps most intriguing is that the solar layers observed by Solanki *et al.* are where the complex action of magnetic-field emergence from the solar interior takes place. Some magnetic-field lines are presumably loaded with hot plasma ions, and will eventually become visible at extreme-ultraviolet and X-ray wavelengths. Solanki *et al.* infer from their analysis that they are indeed looking at emerging magnetic plasma structures. They demonstrate that, in at least one case, their reconstruction of the magnetic-field conditions is detailed enough to enable them to identify a 'current sheet' — the first claimed detection through magnetic-field measurements of such a structure, which marks the divide between regions of different magnetic polarity.

This result is very likely to be the opening salvo in a real revolution in solar research. Theorists will want to know on a quantitative level whether such current sheets are common; whether the heating associated with the dissipation of these currents can account for the heating of the observed plasmas; and whether time sequences can be constructed for magnetic-flux emergence and current-sheet formation (and dissipation). Observers and modellers will want to develop alternative observational strategies and analysis methods to confirm these results.

Perhaps most exciting is the prospective interaction between observers and computational solar physicists. Modern simulation tools are now so sophisticated that quantitative comparison between observation and simulation should be possible. Such comparisons should further our understanding of the mechanism (or mechanisms) that are actually responsible for the heating and structuring of the outer solar atmosphere. Personally, I can't wait for the revolution! ■
Robert Rosner is in the Department of Astronomy

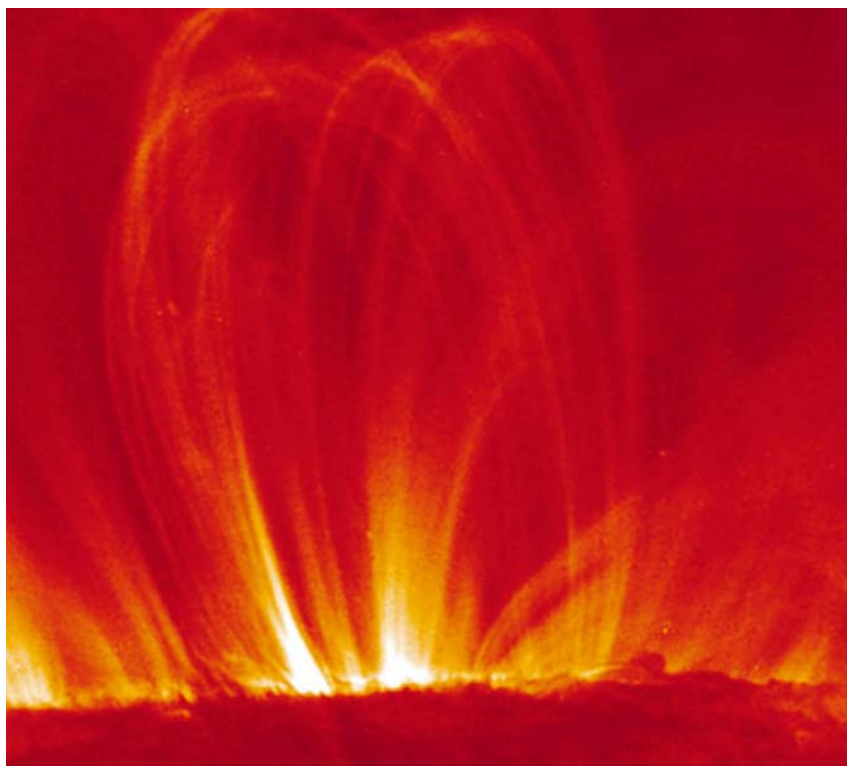


Figure 1 Magnetic moment. This image of a 'quiet' region of the solar corona was taken by the TRACE satellite. Most of the material imaged lies in loop-like structures; the largest loops shown reach heights of around 120,000 km. Temperatures are typically a million degrees kelvin.

and Astrophysics, and the Department of Physics, University of Chicago, and at the Argonne National Laboratory, 9700 South Cass Avenue, Argonne, Illinois 60439, USA.
e-mail: r-rosner@uchicago.edu

1. Solanki, S. K., Lagg, A., Woch, J., Krupp, N. & Collados, M. *Nature* **425**, 692–695 (2003).
2. Rosner, R., Tucker, W. & Vaiana, G. S. *Astrophys. J.* **220**, 643–665 (1978).
3. Parker, E. N. *Astrophys. J.* **330**, 474–479 (1988).
4. Zirin, H. *Astrophysics of the Sun* (Cambridge Univ. Press, 1988).
5. Penn, M. J. & Kuhn, J. R. *Astrophys. J.* **441**, L51–L54 (1995).

Prion diseases

A nucleic-acid accomplice?

Byron Caughey and David A. Kocisko

Prion proteins that trigger a cascade of protein misfolding in the brain are suspected of being the sole transmissible cause of some brain-destroying diseases. But nucleic acids could be their partner in crime.

The seductive notion that a modified host protein might be the sole infectious agent of transmissible spongiform encephalopathies (TSEs) has tantalized the scientific world since the idea was suggested more than 30 years ago¹. The ensuing 'prion hypothesis' was developed following the discovery that a host-cell protein called prion protein (PrP^C) can exist in TSEs in an abnormal form (PrP^{Sc}). Considerable circumstantial evidence is consistent with this Nobel-prizewinning hypothesis, which holds that prions are "transmissible particles that are devoid of nucleic acid"²; however, direct proof of the 'protein-only' nature of TSE agents has remained elusive. In this issue, Deleault *et al.*³

provide striking evidence that nucleic acids might, after all, be involved in propagating TSE infectivity. Writing on page 717, the authors suggest that vertebrate single-stranded RNA is required for the *in vitro* amplification of PrP^{Sc}.

How might prions replicate if they carry no genetic material? The premise of most models of prion replication is that the abnormal form, PrP^{Sc}, propagates itself by inducing a conformational change in PrP^C, and/or triggering its aggregation. Evidence for this view first came from cell-free reactions, in which PrP^{Sc} induced the conversion of PrP^C into a PrP^{Sc}-like aggregated state⁴. Both this conversion product and PrP^{Sc} are partially resistant to destruction by protein-

digesting enzymes (unlike PrP^C), and so they are often called PrPres.

Cell-free conversion reactions are highly specific in ways that reflect biological features of TSE diseases⁵. For example, different TSE agents show particular patterns of interspecies transmission — similar patterns of agent-dependent conversion are also seen in the *in vitro* reactions. But whether PrP^C is converted to the pathological form in the same way in the brain is unknown. Indeed, *in vitro* conversion models suffer from two limitations: PrPres formed in cell-free reactions has never been shown to constitute new TSE infectivity in animals, and most *in vitro* reactions yield substoichiometric quantities of PrPres relative to the PrP^{Sc} initially added⁵. But much higher yields of PrPres are obtained when homogenates from normal brains and TSE-affected brains are mixed together, suggesting that cellular cofactors might be important for PrP^{Sc} propagation *in vivo*⁶. A number of potential cofactors have been proposed, including nucleic acids, which are known to interact with PrP molecules *in vitro*^{7–9}.

Deleault *et al.*³ have now looked at the effect of nucleic acids on the propagation of PrPres in brain homogenates. They found that PrPres was amplified sixfold after mixing infected and normal brain homogenates. But when the homogenates were depleted of single-stranded RNA, no amplification was detected. In contrast, adding high concentrations of exogenous RNA (extracted from uninfected hamster or mouse tissues) boosted PrPres amplification about 24-fold. So these results suggest that single-stranded RNA molecules are necessary for PrPres amplification. Interestingly, the authors found that the activity of the RNA shows some species specificity — RNA from invertebrates failed to support amplification. Further characterization showed that the most active RNA contained more than 300 nucleotides and co-purified with ribosomal RNA (a form of RNA found at the site of protein synthesis in cells). But beyond these general characteristics, the identity of the active RNA is unknown.

Do the new findings refute the popular protein-only prion hypothesis? Not necessarily, because, as the authors point out, RNA might be a host-cell cofactor required for the conversion of PrP^C, rather than an obligatory or informational component of the transmissible prion. Deleault *et al.* showed that active RNA is also present in extracts of normal mammalian tissues, so these molecules are clearly not specific to TSE infections. Furthermore, biochemical analyses of extensively purified PrP^{Sc} preparations have provided no evidence that a nucleic acid component of the size specified by Deleault *et al.* is an essential element of an infectious unit¹⁰. Nonetheless, the possibility remains that small, variable and/or

fragmented host-derived RNA sequences stabilize PrP^{Sc} infectivity.

How might RNA promote the conversion of PrP^C? Nucleic acid molecules (both RNA and DNA) are known to bind to PrP^C, and under certain circumstances DNA can promote conformational changes in PrP^C and trigger its aggregation into fibrils^{7,8}. It is possible that RNA behaves similarly during PrPres amplification. Curiously, Deleault *et al.* saw no effect on PrPres propagation after adding DNA molecules to the reactions or depleting homogenates of endogenous DNA using DNA-digesting enzymes. However, the added DNA molecules were probably far too small to be active, and the effects of the endogenous DNA might have been blocked by bound proteins.

Large, negatively charged nucleic acid molecules (whether RNA or DNA) could act as scaffolds or catalytic templates that promote PrP^{Sc}–PrP^C interactions^{8,11}. But in intact cells, nucleic acids and PrP molecules are usually separated by cell membranes: RNA molecules are found in the nucleus and cytoplasm, whereas PrP^C and PrP^{Sc} molecules are mainly found outside the cell, on the cell surface or in intracellular organelles. Given these apparent logistical difficulties, it is worth considering a different view — that the RNA molecules might act as surrogates for a different cofactor that is crucial for the conversion of PrP^C *in vivo*, but which is in short supply in the *in vitro* reactions.

For instance, sulphated glycosaminoglycans (which, like RNA, are large polyanions) are found in the same parts of the cell as PrP^C and PrP^{Sc}, and these molecules can strongly modulate PrPres formation *in vitro* and *in vivo* (see refs 11, 12 and references therein). Deleault and colleagues found that various other polyanions, including sulphated glycans, had no effect on PrPres amplification. However, given the complexity of brain homogenates, it is possible that the effects of these added polymers were masked by interactions with

vast excesses of molecules other than PrP.

Of all the polyanions that Deleault *et al.* tested, large single-stranded RNA molecules were most efficient at promoting PrPres amplification. The authors' observations raise a number of questions. Can RNA from different species influence the patterns of interspecies transmission of TSE agents, and the propagation of distinct TSE 'strains' and PrP^{Sc} conformations? Does the presence of a specific type of RNA affect TSE infectivity? And does RNA influence the progression of other diseases involving abnormal protein aggregation, such as Alzheimer's, Parkinson's or Huntington's diseases?

Whatever the answers to these questions, the ability of RNA to induce the amplification of PrPres might also have practical implications. It is important to be able to detect very small amounts of PrPres in order to protect the food supply and for early diagnosis of TSEs such as Creutzfeldt–Jakob disease and bovine spongiform encephalopathy (BSE). Using RNA to amplify PrPres in cell-free conversion reactions could improve the sensitivity and reliability of diagnostic tests. So Deleault and colleagues' findings should be of interest whether or not RNA molecules turn out to be important players in PrPres amplification *in vivo*. ■

Byron Caughey and David A. Kocisko are at the NIAID/NIH Rocky Mountain Laboratories, Hamilton, Montana 59840, USA.
e-mail: bcaughey@niaid.nih.gov

1. Griffith, J. S. *Nature* **215**, 1043–1044 (1967).
2. Prusiner, S. B. *Proc. Natl Acad. Sci. USA* **95**, 13363–13383 (1998).
3. Deleault, N. R., Lucassen, R. W. & Supattapone, S. *Nature* **425**, 717–720 (2003).
4. Kocisko, D. A. *et al.* *Nature* **370**, 471–474 (1994).
5. Caughey, B. *et al.* *Adv. Prot. Chem.* **57**, 139–169 (2001).
6. Saborio, G. P., Permanne, B. & Soto, C. *Nature* **411**, 810–813 (2001).
7. Nandi, P. K., Leclerc, E., Nicole, J. C. & Takahashi, M. *J. Mol. Biol.* **322**, 153–161 (2002).
8. Cordeiro, Y. *et al.* *J. Biol. Chem.* **276**, 49400–49409 (2001).
9. Gabus, C. *et al.* *J. Biol. Chem.* **276**, 19301–19309 (2001).
10. Meyer, N. *et al.* *J. Gen. Virol.* **72**, 37–49 (1991).
11. Wong, C. *et al.* *EMBO J.* **20**, 377–386 (2001).
12. Ben-Zaken, O. *et al.* *J. Biol. Chem.* **278**, 40041–40049 (2003).

Materials science

Zero-expansion plan

Arthur Sleight

Some materials don't expand or contract as they are heated. A new example is a metallic compound, for which the movement of electrons between atoms is the likely explanation for its 'zero thermal expansion'.

Most solids expand on heating, because the distance between their atoms increases. This is known as positive thermal expansion. Some materials, however, shrink when they are heated, which is negative thermal expansion. On page 702 of this issue, Salvador *et al.*¹ describe a compound (of ytterbium, gallium

and germanium, YbGaGe) in which a mechanism for negative thermal expansion compensates for the forces causing the positive effect. As a result, this material shows almost zero thermal expansion. Unlike other materials with this unusual property, YbGaGe is electrically conductive, and Salvador *et al.* propose that its lack of thermal expansion is

due to motion of electrons through the material: electrons move from Yb atoms to Ga atoms as the temperature increases, causing Yb atoms to shrink, but creating a negligible increase in the size of Ga atoms.

If the distance between bonded atoms is plotted against the sum of the attractive and repulsive forces between these atoms, the distinctive shape of an 'anharmonic' potential well is normally the result (Fig. 1). It is the anharmonicity of this well — it's 'lop-sidedness' — that allows interatomic distances to increase with increasing temperature: a harmonic well would produce no thermal expansion or contraction with increasing temperature, no matter how wildly the atoms might vibrate as the temperature rises. If the bonds between the atoms are very strong, the shape of the potential well is close to harmonic, and the material's thermal expansion is low; weak bonds lead to high thermal expansion.

Instances of negative thermal expansion in solids had been considered rare, but many oxides have recently been found to have this property². This behaviour results not from a decrease of interatomic distances between bonded atoms with increasing temperature, but from a decrease in the distance between non-bonded atoms. For example, negative thermal expansion in the oxide $\text{Y}_2\text{W}_3\text{O}_{12}$ is due to the decrease in the Y–W separation as the central oxygen atom vibrates perpendicular to the Y–O–W linkage³. The same is true for CuScO_2 , in which the thermal motion of copper atoms perpendicular to the O–Cu–O linkage similarly results in a decrease of the O–O separation⁴. In some cases, this mechanism just balances the usual forces that produce positive thermal

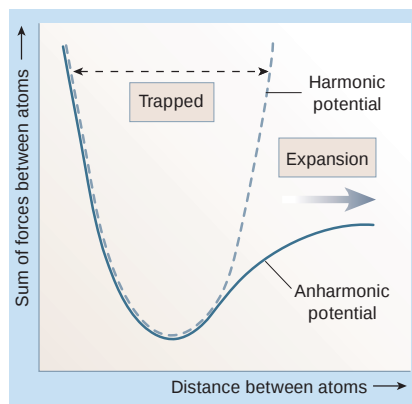


Figure 1 The anharmonic potential well. In many solids, the sum of the forces between bonded atoms over varying interatomic distance describes an anharmonic potential well. Energetically, the preferred state for an atom is at the bottom of the well. But the anharmonic nature of the well means that it is possible for atoms to move further apart when heated — although there is a considerable energetic barrier to them moving closer together. If the potential well were harmonic, expansion would also be blocked.

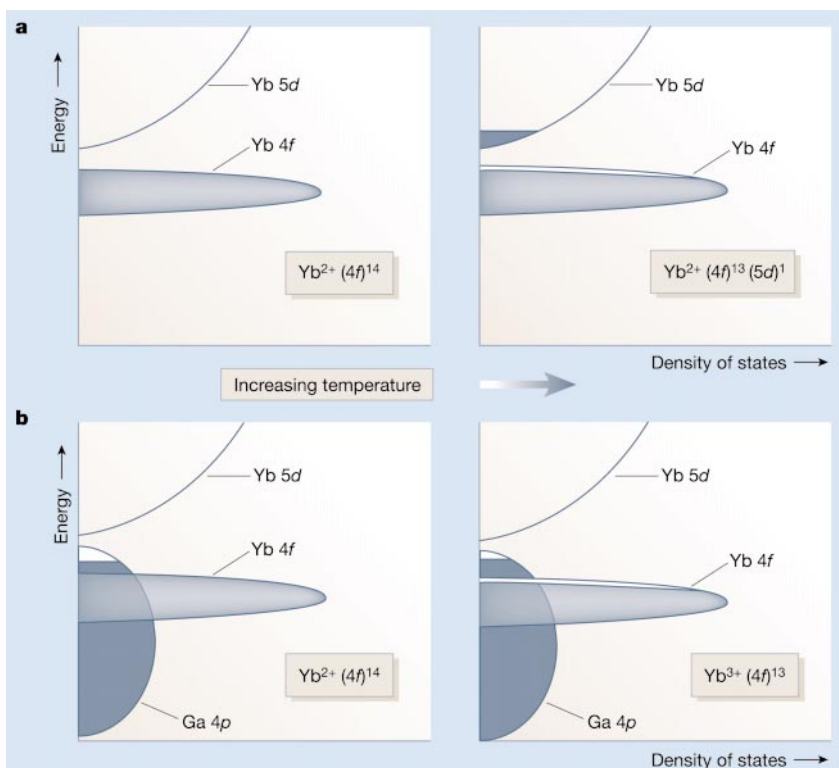


Figure 2 Valence degeneracy and ion size. a, With increasing temperature, electrons are excited from the 4f band to the 5d band in the Yb^{2+} ion (occupied states are indicated by shading). The valency of the Yb ion is unchanged (the overlap of these bands is called valence degeneracy), but the ion shrinks in size. b, There is a similar effect in the material YbGaGe, studied by Salvador *et al.*¹. Now when the temperature is raised, electrons are excited from the Yb 4f band to the Ga 4p band. The Yb^{2+} ion becomes a smaller Yb^{3+} ion. (The 4f band for only one of the two types of Yb in YbGaGe is shown.)

expansion (for example, in TaO_2F and amorphous SiO_2)⁵. Polymers frequently show negative thermal expansion along the chains of strong bonds, but the positive thermal expansion that arises from the weak bonds between chains gives, overall, a large thermal expansion.

For most atoms, their ionization states (such as Na^+ or Fe^{2+}) correspond to a single configuration of electrons in shells around the nucleus. But for certain rare-earth ions, a particular ionization state (or valence state) may exist in two different electronic configurations. In terms of the occupation of the outermost electron orbitals, for Sm^{2+} these configurations are $(4f)^6$ or $(4f)^5(5d)^1$, and for Yb^{2+} they are $(4f)^{14}$ or $(4f)^{13}(5d)^1$ — the electrons may be wholly within the f shell, or one may move into the d shell, but the ionization state is unchanged. Mixed electronic configurations sometimes occur. (Such mixed electronic configurations have sometimes been termed mixed valent states, or intermediate valent states. This is incorrect because the two electronic configurations have the same number of electrons.)

However, the ion with a d-electron is significantly smaller than the ion without a d-electron. When both electronic configurations are present in a material containing rare-earth ions, the relative amount of each

depends on the temperature: if the population with the d-electron increases as the temperature rises (which is frequently the case), there is a corresponding drop in the average size of the ion. This accounts for negative thermal expansion in the compound YbCuAl below 60 K, and in $\text{Sm}_{0.75}\text{Y}_{0.25}\text{S}$ over the wide, measured range of 4.2 to 350 K (refs 6 and 7, respectively). But there are apparently no examples of materials in which this mechanism balances positive thermal expansion to give almost zero thermal expansion over a large temperature range.

The negative thermal expansion of YbGaGe depends on the overlap of the Yb 5d and the Ga 4p bands (Fig. 2). As the temperature increases, the electron density shifts from the Yb 4f band to the Ga 4p band; the Yb ion shrinks in size, but the Ga ion is essentially unchanged. Strong negative thermal expansion along two axes of the hexagonal unit cell of YbGaGe is compensated by the somewhat stronger positive thermal expansion along the third axis, to give a thermal expansion of unit-cell volume that is very close to zero over the temperature range from 100 to 400 K.

It is easy to imagine the usefulness of materials that do not expand or contract appreciably with changing temperature (for instance, in building spacecraft).



100 YEARS AGO

Referring to Mr. Hardy's experiment described in his letter in *Nature*, October 8, it is easy to show that whatever the intensity of radio-activity might be at the surface of the sun, by mere surface ratios and assuming no absorption its activity per unit area at the distance of the earth must fall to about one forty-thousandth part... This supposes no absorption from, possibly, some thousands of miles of solar atmosphere. Moreover, we assume in the comparison a sun of solid radium bromide. It would appear, however, that a very small percentage of this body in the materials of the sun would suffice to account for many millions of years of solar heat. ... The absence of β and γ radiations at the earth's surface is, therefore, not a weighty argument against the presence of radium in the sun. The arguments in favour of supposing that this element exists in the sun are:— (1) The presence of radium on the earth; (2) the high atomic weight of radium; (3) the presence of helium in the sun; (4) Arrhenius's theory of the Aurora Borealis; (5) the fact that the estimate of the duration of solar heat from the dynamical source appears to run counter to geological data. J. Joly
From *Nature* 15 October 1903.

50 YEARS AGO

FORTHCOMING EVENTS

Monday, October 19

INSTITUTE OF ELECTRICAL ENGINEERS (at Savoy Place, London, W.C.2), at 5.30 p.m. — Discussion on "Television".

Tuesday, October 20

UNIVERSITY COLLEGE, LONDON (in the Anatomy Theatre, Gower Street, London, W.C.1), at 1.15 p.m. — Prof. J. B. S. Haldane, F.R.S.: "The Physiology of Diving".

INSTITUTE OF PHYSICS, EDUCATION GROUP (at the Royal Institution, Albemarle Street, London, W.1), at 6.30 p.m. — Prof. O. R. Frisch, F.R.S.: "Atomic Energy — How it all Began".

Wednesday, October 21

GEOLOGICAL SOCIETY OF LONDON (at Burlington House, Piccadilly, London, W.1), at 5 p.m. — Prof. J. Z. Young, F.R.S.: "The Evolution of Vertebrate Organization".

Thursday, October 22

UNIVERSITY COLLEGE, LONDON (in the Anatomy Theatre, Gower Street, London, W.C.1), at 1.15 p.m. — Dr. Randolph Quirk: "Careless Talk: Some Features of Everyday Speech".
From *Nature* 17 October 1953.

Like YbGaGe, cordierite ($\text{Mg}_2\text{Al}_4\text{Si}_5\text{O}_{18}$), β -eucryptite (LiAlSiO_4) and NZP ($\text{NaZr}_2\text{P}_3\text{O}_{12}$) all show anisotropic thermal expansion: they have hexagonal structures in which thermal expansion along two axes is compensated by the opposite sign of thermal expansion along the third. These solids are normally based on polycrystalline aggregates, and as the temperature changes the stresses induced by this anisotropic expansion of the unit cell can cause 'microcracks' to develop, which degrade the mechanical properties of the material. The microcracking problem can sometimes be overcome if the crystallites in the solid are sufficiently small. But microcracking will probably be far less of a problem in a metallic material such as YbGaGe.

The oxides that are known to have negative or very low thermal expansion are all electrical insulators. But for many applications, low-thermal-expansion materials that have high electrical or thermal conductivity

would be desirable. To date, the only candidate material has been 'Invar', an iron–nickel alloy, whose low thermal expansion is related to its magnetic properties⁶. Now, from Salvador *et al.*¹, we have a metallic material whose very low thermal expansion is based on a distinctly different mechanism. ■

Arthur Sleight is in the Department of Chemistry, Oregon State University, Corvallis, Oregon 97331-4003, USA.

e-mail: arthur.sleight@orst.edu

1. Salvador, J. R., Guo, F., Hogan, T. & Kanatzidis, M. G. *Nature* **425**, 702–705 (2003).
2. Sleight, A. W. *Inorg. Chem.* **37**, 2854–2860 (1998).
3. Forster, P. W. & Sleight, A. W. *Int. J. Inorg. Mater.* **1**, 123–127 (1999).
4. Li, J., Yokochi, A., Amos, T. G. & Sleight, A. W. *Chem. Mater.* **14**, 2602–2606 (2002).
5. Tao, J. Z. & Sleight, A. W. *Solid State Chem.* **173**, 45–48 (2003).
6. Mattens, W. C. M., Hölscher, H., Tuin, G. J. M., Moleman, A. C. & de Boer, F. R. *J. Magn. Mag. Mater.* **15–18**, 982–984 (1980).
7. Mook, H. A. & Holtzberg, F. in *Valence Fluctuations in Solids* (eds Falicov, L. M., Hanke, W. & Maple, M. B.) 113–118 (North-Holland, Amsterdam, 1981).
8. van Schilfgaarde, M., Abrikosov, I. A. & Johansson, B. *Nature* **400**, 46–49 (1999).

Evolution

Opportunity versus innovation

Paul H. Harvey and Andy Purvis

Why have some evolutionary lineages produced many more species than others? As far as one large group of birds is concerned, being in the right place at the right time is a plausible answer.

A tradition in biology has been taxonomy, the classification of organisms into hierarchical groupings: the identification of species, the grouping of species into genera, genera into tribes, tribes into families, and so on. Many biologists have long been preoccupied with going further and attempting to construct phylogenies — evolutionary relationships — from that information. But the reverse procedure of building evolutionary trees from molecular data, and then defining taxonomic groupings and levels by dates of common ancestry, has opened up new avenues for studying evolutionary processes.

One such process is the occurrence of species radiations, in which certain evolutionary lineages have diversified and produced far more species than others. What factors lie behind this phenomenon? As he describes in *Proceedings of the Royal Society*, Robert Ricklefs¹ has tackled the question by looking at data for the passerines. This is a group of birds, at the taxonomic level of order (one above families), which includes over 5,000 species and such well-known examples as crows, thrushes and finches. One common assumption in evolutionary biology is that key morphological or behavioural innovations spark off adaptive radiations. For example, plant-feeding in insects has evolved on several occasions and seems

generally to have resulted in an increased net rate of speciation². Ricklefs concludes, however, that key innovations are unlikely to have been the major cause of variation in speciation rates among the passerines. Instead, those species that happened to be in the right place at the right time to exploit ecological opportunities were the progenitors of the major radiations.

Ricklefs' starting point was an earlier study by Sibley and Ahlquist³. They considered species of birds that last shared a common ancestor between 21 million and 25 million years ago to belong to the same family. Assuming that Sibley and Ahlquist's molecular dating techniques were approximately correct, the number of extant species differs widely among passerine lineages of this age. For example, the most speciose family of passerine birds contains 993 species whereas several others contain just one. If speciation and extinction rates did not differ among lineages, chance alone would ensure considerable variation in the number of species per family; at any point in time the distribution of species among families would follow a geometric probability distribution⁴. But when compared with such a distribution, the variance in number of species in passerine families is far too high: several families have too many species and too many families have very few.

Astronomy

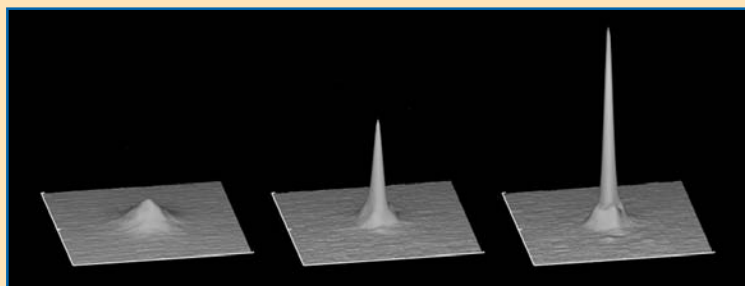
Faking it

Astronomers are celebrating 'first light' at the newly adapted Keck II 10-metre telescope in Hawaii. Keck II now has a laser guide-star system — the first on such a large telescope. By creating an artificial star in the sky, it will greatly improve the telescope's ability to image distant galaxies.

The system is based on adaptive optics (AO), an established technique by which images taken at Earth-bound telescopes can be corrected for the blurring caused by the planet's atmosphere. Using AO produces much sharper images. But the scheme usually relies on there

being a bright guide star in the vicinity of the object under observation, against which the correction is calibrated.

At Keck II, a 15-watt laser beam is fired into the sky, creating a glowing patch in a natural layer of sodium atoms, 90 km above the Earth's surface. Using this glow as an artificial guide star, the images from the telescope become sharper still — as shown in the picture here, of the Strehl ratio



(a measure of image 'perfection') for a star without AO, with AO and with AO plus the laser guide star (left to right).

But it's not just a question of

resolution. The laser system can also be directed at any region of sky: astronomers need no longer rely on good fortune to find a guiding star.

Alison Wright

That is not surprising. What is surprising, biologically, is that neither species-rich nor species-poor families are phylogenetically clustered: that is, if a passerine bird belongs to a single-species family, the most closely related family is no more likely to be species-poor than any other family. If morphological or behavioural innovations were the cause of high speciation or low extinction rates, we should surely expect closely related families, whose species are more similar than those from randomly selected families, to contain similar numbers of species. In contrast, closely related groups of flowering plants do, indeed, have similar numbers of extant species⁵. Furthermore, the passerine families have more than three times as many species, on average, as do the bird families not in Ricklefs' study. The point here is that *within* the passerines there seem to be no strong phylogenetic associations with net speciation rates.

Perhaps the timing over which radiations occur within passerine birds is detectable only among more closely related clades than families (a clade consists of all the species sharing a single common ancestor). The rank of tribe is the next major taxonomic level down, and species belonging to the same passerine tribe last had a common ancestor 10–16 million years ago. But when Ricklefs searched for a phylogenetic pattern of species richness among tribes, he again drew a blank. The families with more species also contained more tribes although, interestingly, those tribes were not unusually speciose. His suggestion, then, is that the events that resulted in speciose families occurred sometime after the origin of families (21–25 million years ago) but before the origin of tribes (10–16 million years ago).

By concentrating on the few families containing six or more tribes (the average

number of tribes per family is about 2.3), Ricklefs proposes that unusual expansion of geographical ranges might explain the unusually species-rich clades. One example is the invasion of South America by North American fringillid finches. Another is the spread of the corvid family — crows and ravens — from Australasia as tectonic plate movement brought Asia close by.

But what about species-poor taxa? Here the explanation is that tribes with few species simply never had a chance to radiate because, for example, they are restricted to remote locations away from the continental land-masses or are dietary specialists. As for the majority of families and tribes with more average numbers of species, a constant-rates speciation–extinction process seems to model their distributions perfectly well with little need to seek key innovations.

But the argument that data fit a model except when they do not fit a model must be viewed with suspicion — especially when, as in one analysis, fewer than half do fit. What is more, there are some weak but significant correlations between species richness and mating system or sexual dimorphism among passerine birds, which need to be explained⁶. Tribes in which mate choice by females is a strong selective force — often leading to the evolution of brightly coloured males — tend to be rich in species. The most straightforward explanation of this effect of sexual selection is that, when female choice is strong, separated populations tend to diverge rapidly in both male plumage and female preference, accelerating speciation.

Ricklefs suggests either that those characters that affect speciation are phylogenetically clustered at taxonomic levels higher than the family, and so cannot be revealed by his analyses, or that 'reversed causation' was involved: high species richness within a clade

somehow promoted sexual selection. A third reason, of course, is 'unreversed causation': sexual selection may, indeed, have promoted speciation (or reduced extinction), but Ricklefs' analyses have not detected this weak effect. Further work should be able to distinguish among these explanations, perhaps by combining Ricklefs' taxon-based approach with trait-based⁶ analyses (testing explicitly whether particular characteristics are repeatedly associated with high diversity) and tree-based methods⁷ (where the structure and shape of the phylogeny are analysed for clues about how it grew).

Whatever the outcome, Ricklefs has provided a fresh perspective from which evolutionary biologists, ecologists and ornithologists can better understand the diversification of the passerines. And, importantly, he has provided a blueprint for similar analyses of other taxa when dated phylogenies that link extant species become available.

Paul H. Harvey is in the Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK.

e-mail: paul.harvey@zoo.ox.ac.uk

Andy Purvis is in the Department of Biological Sciences, Imperial College London, Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK.

e-mail: a.purvis@imperial.ac.uk

1. Ricklefs, R. E. *Proc. R. Soc. Lond. B* doi:10.1098/rspb2003.2489 (2003).
2. Mitter, C., Farrell, B. & Wiegmann, B. *Am. Nat.* **132**, 107–128 (1988).
3. Sibley, C. G. & Ahlquist, J. E. *Phylogeny and Classification of Birds of the World* (Yale Univ. Press, New Haven, Connecticut, 1990).
4. Kendall, D. G. *Biometrika* **35**, 6–15 (1948).
5. Magallón, S. & Sanderson, M. J. *Evolution* **55**, 1762–1780 (2001).
6. Barraclough, T. G., Harvey, P. H. & Nee, S. *Proc. R. Soc. Lond. B* **259**, 211–215 (1995).
7. Purvis, A., Nee, S. & Harvey, P. H. *Proc. R. Soc. Lond. B* **260**, 329–333 (1995).

Tissue repair

Gel brought to heal in mice

Curr. Biol. **13**, 1697–1703 (2003)

A new bio-active gel speeds up wound healing in mice. The treatment targets connexin 43 (Cx43), a protein that is involved in cell-to-cell communication and which is overproduced at injury sites.

Cindy Qiu *et al.* tested the gel on mice with open skin wounds. The active ingredients — so-called antisense oligonucleotides — are nucleic acid sequences that prevent cells from making Cx43, by binding to the protein-encoding RNA. A single application enhanced the normal decline of Cx43 protein levels in and around the wound. The number of neutrophils — damaging inflammatory cells that gobble up cellular debris and bacteria — was reduced. Wounds were less swollen and healed more quickly.

The antisense oligonucleotides break down rapidly within the tissues they penetrate, so the effects are transient. The gel kick-starts the healing process, the authors say, and its effects persist long after the oligonucleotides are gone. After 12 days, the scars were smaller, narrower and flatter against the skin.

The gel may prove to be an effective and safe therapy for a variety of wounds, the authors speculate.

Helen R. Pilcher

Vision

In the eye of the beholder

J. Exp. Biol. **206**, 3963–3977 (2003)

Cells in the retina of male flies are specialized for spotting small, fast-moving objects, giving them an edge over their quarry when pursuing females with mating

in mind. The 'lovespot' in the male's eye can detect another fast-moving fly from 76 cm away: female eyes can manage only 33 cm, Brian G. Burton and Simon B. Laughlin have found.

The lovespot's receptors fire strongly in response to small objects, and they respond more quickly, pinpointing the target's position with greater precision, than do females' receptors. They also stop responding more quickly. This reduces blurring, as it erases any after-image from the retina.

There are plenty of examples of eye adaptation for particular purposes. Frogs' eyes, for instance, respond strongly to anything that might be a flying meal. But in this case the relevant events occur further along the image-processing path, not in the receptor cells themselves. Burton and Laughlin's work involved using a small moving light to mimic a fly target for another fly, and recording the electrical output from single receptor cells in the retina.

John Whitfield

Fluid dynamics

Jet-bending explained

Phys. Fluids **15**, 3568–3571 (2003)

Last year, James M. Chwalek *et al.* reported that liquid jets squirting through a nozzle can be deflected by a temperature gradient (*Phys. Fluids* **14**, L37–L40; 2002). Michael P. Brenner and Srinivas Paruchuri have now furnished a simple theoretical explanation for this effect. This kind of understanding is essential for putting the effect to practical use, as it identifies the key factors influencing the jet's trajectory.

Controlling the flow of a liquid jet is central to the function of, for example, ink-jet printers. In some printers it is achieved by using a cumbersome array of

electrodes to deflect individual, electrically charged droplets after the jet has broken up.

Deflecting the unbroken jet would potentially be much cleaner. It might also be useful for fibre spinning of polymer fibres, where the liquid polymer is forced through a nozzle.

Chwalek *et al.* produced deflection angles of several degrees by heating the nozzle asymmetrically. Brenner and Paruchuri show that the main cause is the Marangoni effect — the temperature dependence of surface tension, which sets up a bending stress on a cylindrical jet.

Philip Ball

Neurobiology

Alzheimer's enzyme target

Mol. Cell **12**, 553–563 (2003)

Researchers have struggled to work out how fragments of amyloid- β , which accumulate into plaques in the brains of people with Alzheimer's disease, kill neurons and cause dementia. Sungmin Song *et al.* have now identified a way in which the fragments block the breakdown of certain unwanted neuronal proteins.

Using a gene chip, the team found that amyloid- β ramps up production of an enzyme, E2-25K/Hip-2, which marks redundant proteins for destruction in a cell compartment called the proteasome. It does this by adding a ubiquitin chemical tag.

Levels of E2-25K/Hip-2 are higher than normal in the brains of Alzheimer's patients and in neurons exposed to toxic amyloid- β . Excess enzyme is implicated in amyloid- β -induced neuron death, but the exact mechanism involved is unclear. It causes a build-up of tagged proteins in neurons, and activates two proteins (ASK1 and JNK) that trigger cell death. The enzyme may also tag a protein, UBB+1, that inhibits the proteasome.

The authors speculate that drugs that block E2-25K/Hip-2 might help to prevent some of the symptoms of Alzheimer's disease. But the basic biology would need to be worked out far more before help could be offered to the estimated 15 million sufferers around the world.

Helen Pearson

Reproduction

Being compatible

Genes Dev. **17**, 2502–2507 (2003)

External fertilization can be a hit-and-miss affair. Sperm and egg must first find each other and then unite — a tricky process for marine animals that release their gametes into water. The sea urchin, however, has evolved a series of steps to help the process and to prevent sperm from one species fertilizing the eggs of another.

The sperm express a protein, called bindin, at the tip. This protein helps sperm bind to egg, and it differs between sea urchin species. Its discovery prompted researchers to muse that complementary species-specific egg proteins, or receptors, must also exist.

Now these elusive egg proteins have been identified. Noriko Kamei and Charles G. Glabe looked for species-specific sequences in DNA from ovaries of the sea urchins *Strongylocentrotus franciscanus* and *S. purpuratus*. One sequence that differed between the two species turned out to encode a bindin receptor on eggs, which they dubbed EBR1.

The discovery may shed light on how the species have evolved and help reveal, at the molecular level, how sperm and egg interact to form fertile offspring.

Helen R. Pilcher



Eyes right — male flies have the edge in the mating game.

Single-gene speciation by left–right reversal

A land-snail species of polyphyletic origin results from chirality constraints on mating.

A single gene gives rise to the mirror-image form of a snail's body plan, which could become established as a different species if mating is prevented between snails of different chirality by genital mismatch^{1–3}. Here we use molecular phylogeny to demonstrate the parallel evolution of reversal between left and right lineages of the Japanese land snail *Euhadra*. We find that the different mirror-image forms have evolved in favour of the genetically dominant handedness as a result of single-gene speciation.

Speciation accompanying the left–right reversal of the entire ontogeny (chiral speciation) is unique to snails and can be visualized by the coiling direction of their shells⁴ (Fig. 1a). The chirality (or 'handedness' — occurring as 'sinister' and 'dexter' forms) of snails is determined by the maternal nuclear genotype at a single locus⁵. Because of the physical difficulty of two-way copulation between snails that have opposite coils (Fig. 1b), frequency-dependent selection occurs to eliminate the chiral minority^{3,6–8} and antagonize chiral speciation. Offspring from each mother all develop with an identical chiral phenotype, regardless of their genotype. This peculiar mode of inheritance could be advantageous in enabling the dominant reversal allele to proliferate in small, isolated populations^{2,9}. Once the mirror-image variant exceeds 50%, selection fixes the reversal. Reproductive isolation by handedness would then be complete.

In *Euhadra*, 4 of 20 species are sinistral. We discovered three notable processes of chiral evolution, based on analysis of mitochondrial DNA phylogeny (see supplementary information). First, the sinistral taxa are all derived from a single sinistral ancestor. Second, reversal to a dextral species from the sinistral ancestor has occurred in at least three independent lineages. Third, all of the haplotypes of a dexter *E. aomoriensis* are included within a sinistral *E. quaesita*.

This pattern cannot be explained by introgression of mitochondrial DNA between *E. quaesita* and other dextral species, because opposite handedness prevents hybridization. Both mitochondrial-DNA phylogeny and a similar shell morphology (Fig. 1a) indicate that *E. aomoriensis* is derived from *E. quaesita*. *E. aomoriensis* is not merely a dextral morph of *E. quaesita*, as the two differ in shell sculpture although they are partly sympatric. We conclude that *E. aomoriensis* has speciated from *E. quaesita* by virtue of its chirality. The two species mutually failed in interchiral copulation despite frequent attempts, as did the chiral

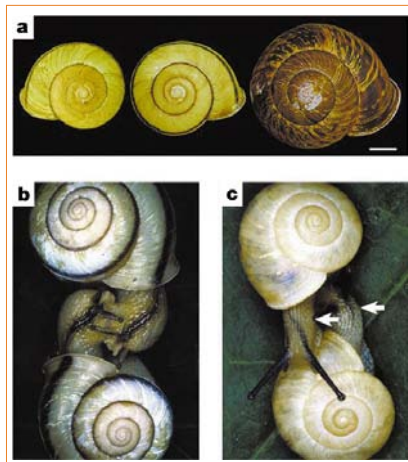


Figure 1 Mirror-image species and copulation in snails. **a**, Shells of *Euhadra*: left, *E. quaesita* (sinistral, or left-handed); middle, *E. aomoriensis* (dextral, or right-handed); right, *E. senckenbergiana* (dextral). Scale bar, 10 mm. The remarkable similarity of the shells reflects the chiral speciation of *E. aomoriensis* from *E. quaesita*, although *E. aomoriensis* has traditionally been considered to be a subspecies of *E. senckenbergiana* (see text). **b**, Successful two-way copulation between dexters of *E. congenita*. **c**, Genital mismatch between a sinistral variant (top) and an ordinary dexter (bottom) of *Bradybaena similaris*. Arrows indicate genital openings that cannot be joined.

variants of *Bradybaena similaris*³ (see supplementary information). Chiral reversal is therefore an acute pre-mating mechanism of isolation.

Reversal to a dextral species has occurred multiple times after only sinistral evolution, so chiral speciation must have been easier to the dexter than to the sinister. We tested chiral inheritance in the family Bradybaenidae, which includes *Euhadra*, using a single dextral adult *Bradybaena similaris* collected with an exceptional sinistral variant³. The dexter, which presumably had been inseminated by multiple conspecific dexters, produced sinistral offspring.

According to the single-locus, delayed-inheritance model with dextral dominance, the dextral mother would have been a recessive homozygote and the sinistral offspring (F_1) would have been heterozygotes. Seven pairs of F_1 sinisters produced 131 dexters (F_2), in which only 30 produced sinistral offspring and others did not. The 30 dexters must therefore have been recessive homozygotes. The ratio of 30 to 101 is consistent with a prediction of allelic segregation at a single locus, with dextral dominance in F_2 ($\chi^2 = 3.01$, $P = 0.62$). The repeated twofold reversals to dextral species must therefore reflect the greater

chance of fixation for the dominant allele, as predicted by theory^{9,10}.

The polyphyly of *E. aomoriensis* in mitochondrial-DNA phylogeny could have resulted from parallel evolution of *E. aomoriensis*, or from introgression or ancestral polymorphism of mitochondrial DNA. As the parallel evolution of three dextral lineages has occurred, multiple peripheral populations of *E. quaesita* may also have fixed the dextral allele independently. The polyphyly of ancestral species is expected where prompt speciation recurs in peripheral isolates¹¹. Incipient single-gene speciation without genome-wide changes must allow *E. aomoriensis* to interbreed with polyphyletic descendants. *E. aomoriensis* may therefore be a unique example of a polyphyletic animal species.

Introgression is impossible unless the reversal allele has persisted in either species under stringent selection. Any introgression must therefore have been limited to a short time before chiral fixation. Ancestral polymorphism would not easily survive chiral speciation in small, isolated populations.

Left–right reversal in reciprocally mating snails can accomplish pre-mating isolation at a single locus, overcoming the majority rule of handedness. Our results indicate that single-gene speciation is possible, at least in hermaphroditic snails, contrary to the traditionally held view^{2,12,13}.

Rei Ueshima*†, Takahiro Asami†‡

*Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo 113-0033, Japan

‡Department of Biology, Shinshu University, Matsumoto 360-8621, Japan

e-mail: asami199@gipac.shinshu-u.ac.jp

†PRESTO, Japan Science and Technology, Kawaguchi 332-0012, Japan

- Gittenberger, E. *Evolution* **42**, 826–828 (1988).
- Orr, H. A. *Evolution* **45**, 764–769 (1991).
- Asami, T., Cowie, R. H. & Ohbayashi, K. *Am. Nat.* **152**, 225–236 (1998).
- Vermeij, G. J. *Nature* **254**, 419–420 (1975).
- Freeman, G. & Lundelius, J. *Wilhelm Roux's Arch. Dev. Biol.* **191**, 69–83 (1982).
- Lipton, C. S. & Murray, J. *Malacologia* **19**, 129–146 (1979).
- Johnson, M. S. *Heredity* **49**, 145–151 (1982).
- Johnson, M. S., Clarke, B. & Murray, J. *Evolution* **44**, 459–464 (1990).
- Van Batenburg, F. H. D. & Gittenberger, E. *Heredity* **76**, 278–286 (1996).
- Asami, T. *Forma* **8**, 263–276 (1993).
- Harrison, R. G. in *Endless Forms: Species and Speciation* (eds Howard, D. J. & Berlocher, S. H.) 19–31 (Oxford Univ. Press, Oxford, 1998).
- Dobzhansky, T. *Genetics and the Origin of Species* (Columbia Univ. Press, New York, 1937).
- Muller, H. J. *Biol. Symp.* **6**, 71–125 (1942).

Supplementary information accompanies this communication on Nature's website.

Competing financial interests: declared none.

Kinetics

Gaussian statistics in granular gases

A kinetic theory^{1,2} of granular media, analogous to that described by Maxwell-Boltzmann statistics for molecular gases and liquids^{3,4}, has long been sought in the quest for a continuum theory. Here we present results from a two-layer, vertically vibrated granular system, in which an oscillating plate drives a horizontal layer of heavy grains, which in turn drives an overlying horizontal layer of lighter grains. In the second layer, but not in the first, we find that the velocity distributions are gaussian — a result that is observed for a wide variety of driving and density parameters, indicating that the second layer's behaviour is much more gas-like than was previously seen in driven granular media. This similarity to molecular gases brings a kinetic theory of granular media a step closer.

Attempts to apply a kinetic theory to granular media have been undermined by their observed non-gaussian velocity distributions^{5–12}, which are inconsistent with Maxwell-Boltzmann statistics. We have designed a set-up (Fig. 1a) in which a two-layer granular medium is shaken vertically on a horizontal plate. The plate moves as

$x(t) = A \times \sin(2\pi ft)$, which corresponds to a dimensionless acceleration of $\Gamma = (2\pi f)^2/g$ where f is the oscillation frequency, A is the oscillation amplitude and g is the acceleration due to gravity. The plate's circular sidewall is topped by a Plexiglas plate coated with a conducting film to prevent electrostatic build-up.

The granular medium is composed of two types of grain. The first layer above the oscillating plate is made up of dimeric 'grains', each consisting of two 3.2-mm steel spheres connected by a metal rod made from a cut bank chain (as used by banks to prevent patrons from removing their pens); each dimer has a mass (m_d) of 185 ± 7 mg. The second layer is composed of monomeric grains that are plastic (Delrin) spheres of diameter 3.175 mm and mass (m_m) 22.9 ± 0.2 mg. The difference in mass between the grains of the first and second layers keeps the dimers and monomers separated, as the monomers cannot displace the heavier dimers.

The system is prepared by building a single layer of dimers on the plate (Fig. 1b) and placing a second layer of Delrin spheres on top (Fig. 1c). We define the coverage, c , of the layer to be the proportion of the maximum number of grains that could theoretically be packed into the area if wall effects are ignored. We collected data at $c = 0.4, 0.6$ and

0.8 , $f = 50, 70$ and 90 Hz, and at dimensionless accelerations (Γ) of 1.75, 2.00 and 2.25.

Velocities in the first and second layers can be measured simultaneously by high-speed digital photography, although measurement by this method in the first layer becomes difficult at high coverage. The two layers show very different velocity distributions (Fig. 1d): in the first, the distribution tails are non-gaussian (typical deviation from gaussian flatness is about 1.5, which is consistent with earlier results^{5–12}), whereas in the second layer the distribution is much more gaussian (typical deviation from gaussian flatness is less than 0.1).

The gaussian-velocity distributions in the second layer indicate that the grains are more 'gas-like' in their behaviour than has previously been found in vibrated granular media. This is probably due to the way in which energy is transferred to the second layer through collisions with the first layer throughout the shaking cycle. By contrast with earlier experiments, the gaussian velocity distribution of the second layer is observed at all coverages, frequencies and accelerations. Our system may therefore be ideally suited to developing a kinetic theory of granular gases.

G. W. Baxter*, J. S. Olafsen†

*School of Science, Penn State Erie, The Behrend College, Erie, Pennsylvania 16563, USA

†Department of Physics and Astronomy, University of Kansas, Lawrence, Kansas 66045, USA

e-mail: jolafsen@ku.edu

1. Haff, P. K. *J. Fluid Mech.* **134**, 401–430 (1983).
2. Grossman, E. L., Zhou, T. & Ben-Naim, E. *Phys. Rev. E* **55**, 4200–4206 (1997).
3. Huang, K. *Statistical Mechanics* (Addison-Wesley, Boston, 1963).
4. Batchelor, G. K. (ed.) *An Introduction to Fluid Dynamics* (Cambridge Univ. Press, Cambridge, 1983).
5. Olafsen, J. S. & Urbach, J. S. *Phys. Rev. Lett.* **81**, 4369–4372 (1998).
6. Olafsen, J. S. & Urbach, J. S. *Phys. Rev. E* **60**, R2468–R2471 (1999).
7. Losert, W., Cooper, D. G. W., Delour, J., Kudrolli, A. & Gollub, J. P. *Chaos* **9**, 682–690 (1999).
8. Kudrolli, A., Wolpert, M. & Gollub, J. P. *Phys. Rev. Lett.* **78**, 1383–1386 (1997).
9. Kudrolli, A. & Henry, J. *Phys. Rev. E* **62**, R1489–R1492 (2000).
10. Blair, D. L. & Kudrolli, A. *Phys. Rev. E* **64**, R050301 1–4 (2001).
11. Rouyer, F. & Menon, N. *Phys. Rev. Lett.* **85**, 3676–3679 (2000).
12. Atwell, J. & Olafsen, J. S. *Phys. Rev. E* (submitted).

Competing financial interests: declared none.

corrigendum

A cloned horse born to its dam twin

Cesare Galli, Irina Lagutina, Gabriella Crotti, Silvia Colleoni, Paola Turini, Nunzia Ponderato, Roberto Duchi, Giovanna Lazzari
Nature **424**, 635 (2003)

It has been drawn to our attention that a successful pregnancy in a goat carrying a genetically identical conceptus has previously been reported¹; complete immune compatibility was demonstrated between the mother and her kid twin (both generated by splitting a single embryo). These findings also indicate that a maternal immune response is not necessary to support a healthy pregnancy.

1. Oppenheim, S. M., Moyer, A. L., Bondurant, R. H., Rowe, J. D. & Anderson, G. B. *Theriogenology* **54**, 629–639 (2000).

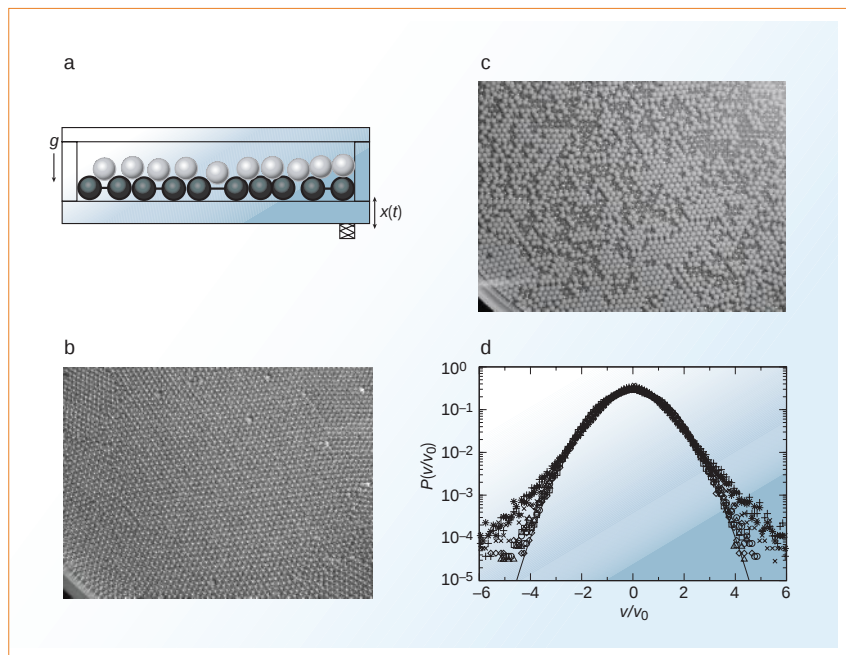


Figure 1 The two-layered granular system. **a**, Geometry of the experimental set-up; the accelerometer is shown at the bottom, on the right. **b**, Photograph of the dimeric layer: diameter of spheres, 3.2 mm; dimer mass, 185 mg. Note the extensive ordered region. **c**, Photograph of white Delrin balls (coverage (c) = 0.4) at rest on top of the dimer layer (roughly one-sixth of the cell is shown). **d**, Velocity statistics for the first- and second-layer grains, for dimensionless acceleration $\Gamma = 2.0$. The distribution of the first-layer grains is non-gaussian. First layer: crosses, $f = 70$ Hz, $c = 0.40$; plus symbols, $f = 50$ Hz, $c = 0.60$; asterisks, $f = 90$ Hz, $c = 0.60$. For the second-layer grains, the renormalized distribution is independent of system parameters within the range studied and is nearly gaussian (solid line) in shape. Second layer: diamonds, $f = 70$ Hz, $c = 0.40$; squares, $f = 50$ Hz, $c = 0.60$; circles, $f = 90$ Hz, $c = 0.60$; triangles, $f = 70$ Hz, $c = 0.80$.

Planning for smallpox outbreaks

Neil M. Ferguson¹, Matt J. Keeling², W. John Edmunds³, Raymond Gani⁴, Bryan T. Grenfell⁵, Roy M. Anderson¹ & Steve Leach⁴

¹Department of Infectious Disease Epidemiology, Faculty of Medicine, Imperial College London, St Mary's Campus, Norfolk Place, London, W2 1PG, UK

²Department of Biological Sciences and Mathematics Institute, University of Warwick, Gibbet Hill Road, Coventry, CV4 7AL, UK

³Health Protection Agency, CDSC, 61 Colindale Avenue, London, NW9 5EQ, UK

⁴Health Protection Agency, CAMR, Porton Down, Salisbury, Wiltshire SP4 0JG, UK

⁵Department of Zoology, University of Cambridge, Downing Street, Cambridge, CB2 3EJ, UK

Mathematical models of viral transmission and control are important tools for assessing the threat posed by deliberate release of the smallpox virus and the best means of containing an outbreak. Models must balance biological realism against limitations of knowledge, and uncertainties need to be accurately communicated to policy-makers. Smallpox poses the particular challenge that key biological, social and spatial factors affecting disease spread in contemporary populations must be elucidated largely from historical studies undertaken before disease eradication in 1979. We review the use of models in smallpox planning within the broader epidemiological context set by recent outbreaks of both novel and re-emerging pathogens.

Events of recent years have heightened awareness of the potential threat of bioterrorism¹, with smallpox considered to pose the greatest risk owing to the lethality (around 30%, depending on age and other factors^{2,3}) and transmissibility of the virus. Although once endemic in many human populations, smallpox was eradicated in 1979 largely as a result of mass vaccination reinforced by other highly focused control measures³. Today, although viral samples are officially retained in only two locations, the existence of other sources cannot be ruled out⁴. In the face of such a difficult to quantify, unlikely, but potentially serious threat, contingency planning demands a rational assessment of the scale of casualties that a smallpox attack might cause, and identification of what controls might be optimal in minimizing its effects^{5–8}. The latter task is complicated by the severe adverse effects of vaccination experienced by a significant minority of individuals^{9,10}, such that a national mass vaccination campaign could cause more deaths than an isolated smallpox epidemic. A second key problem is the passage of time since the last smallpox outbreak: human populations, mobility and patterns of social interactions have changed in the last 30 years, complicating extrapolation from historical epidemics to the prediction of future outbreaks.

Mathematical models of the transmission of infectious agents are valuable tools in making such assessments, because they can integrate epidemiological and biological data to give quantitative insights into patterns of disease spread and the effect of interventions. Examples include the design and evaluation of childhood disease immunization programmes^{11,12}, predicting the demographic impact of the HIV epidemic in different regions¹³, and analysing the spread and control of the 2001 foot-and-mouth epidemic in Britain^{14–17}.

Following this philosophy, four recently published analyses of the potential spread of smallpox virus in modern urban communities all rely heavily on mathematical modelling^{5–8}. Given that a key aim of these studies was to inform public health planning, it is unfortunate that their conclusions differ as to which type of vaccination strategy might be optimal, and on the scale of casualties likely to result from a smallpox attack. As mathematical modelling is the only way that we can examine the possible impact of different release and control scenarios, it is important to understand the strengths and weaknesses of different modelling approaches and how model assumptions affect the conclusions drawn.

Epidemic dynamics

At its simplest, an epidemic is a chain reaction of disease spread within a population (Box 1). The growth of an epidemic is

principally governed by two factors: the number of secondary cases generated by one primary case at the start of the epidemic, termed the basic reproduction number or R_0 , and the average time taken for the secondary cases to be infected by a primary case, termed the generation time or T_G (ref. 18). R_0 essentially determines how intensive a policy will need to be to control the epidemic, whereas both T_G and R_0 determine the time available to implement suitably intensive controls.

Thus, control policies for diseases which are highly infectious with short incubation periods, such as measles ($R_0 \approx 17$, $T_G \approx 11$ days), tend to focus on the long-term reduction of the recruitment of susceptible people, through widespread childhood immunization^{11,12,19}. In contrast, smallpox is both less infectious ($R_0 = 4–10$; refs 20, 21) and has a much longer incubation period ($T_G \approx 21$ days); thus, if an outbreak is detected in its earliest stages, there is sufficient time for localized control measures to be adopted. The feasibility of localized outbreak control during the final phase of the global smallpox vaccination programme, once a degree of 'herd immunity' had been created, is a key reason why smallpox was the first major viral pathogen to be eradicated worldwide^{3,6}.

Evaluating the effectiveness of a potential control policy is not straightforward. A range of key epidemiological and social processes can have significant effects on the likely success of any given intervention. Greater model realism in describing disease biology and human contact patterns is needed to understand which controls work best under different conditions.

Increasing model realism

The simplest description of how the above processes and parameters govern disease transmission is represented mathematically by a crude model¹⁸ (Box 1), which splits the population into 'susceptible', 'infected' and 'recovered' categories. Given its simplicity, this SIR model has been remarkable for the qualitative insights it has given into the epidemiology of a wide range of pathogens^{18,22}. However, quantitative predictions that can be used in policy formulation, cost-benefit analysis or risk assessment frequently necessitate refinement of the basic model (Box 2), to allow for the inclusion of appropriate biological and behavioural factors.

Fundamental to the SIR model are assumptions that all susceptible people in the population are equally at risk of infection from any infected individual (that is, homogeneous mixing), and that all infected individuals have a constant and equal infectiousness. The latter assumption is clearly invalid for smallpox, where an uninfected incubation period of about 12 days is typically followed by a 2–4-day prodromal period associated with mild symptoms and low infectiousness, then a highly infectious symptomatic period of

about 9 days (Box 2). Inclusion of these disease stages within models of smallpox is essential if the effect of vaccination (or other control measures, such as quarantine) on infectiousness or mortality is to be represented realistically^{5–8}.

Assuming homogeneous mixing of the population is also unrealistic—individuals tend to make contact with household members, workplace colleagues and friends at a much higher rate than random strangers, and such regular contacts will also tend to be in the same geographic vicinity. Hence socio-spatial structure has an important impact on transmission dynamics²³ (see Box 2), particularly for diseases such as smallpox where close contact (for example,

of the type more likely to occur in households and hospitals³) is usually required for transmission. However, incorporation of such detail into models considerably increases their complexity and the number of model parameters⁷ that need to be estimated.

Assessing effectiveness of control strategies

A variety of methods exist for controlling the spread of smallpox, ranging from different vaccination strategies to movement/contact restrictions placed on infectious cases and their contacts (Table 1). Thus a key aspect of policy-orientated epidemic modelling is to assess both the adequacy of current policy and how it might further be optimized. Optimality is principally the minimization of mortality and morbidity, so it is critical that models accurately incorporate expected adverse event rates from vaccination. However, the SARS virus has shown that the economic costs of an epidemic can be out of all proportion to the numbers infected, indicating that minimizing the duration of a smallpox outbreak might also be a critical priority when formulating control strategy.

In all cases, it is critical that models explicitly capture the underlying mechanism of the control policy being investigated. The net effect of a control policy on disease transmission is well characterized by its impact on R : the effective reproduction number during the epidemic (Box 1). However, estimating this effect on R without explicit modelling of the details of a control policy is practically impossible for complex models that include population heterogeneity and a realistic description of disease biology. Models which just assume a priori the effect of a control measure on R (see for example refs 5, 8) have little inferential power because the predicted effectiveness is determined by the assumed value of R .

In addition to R , other factors determine the effectiveness and likely success of any given control measure. These include the likely scale and geographic extent of any bioterrorist attack, associated risks or fatalities due to the control, the disruption to civic society, the level of vaccine uptake, overall resource requirements, and the ability of health agencies to implement policies. Models should therefore incorporate realistic logistical constraints on policy implementation⁶, and, if needed, economic costs. There may also be a subtle interplay with epidemic dynamics here. For example, ring vaccination is constrained to the speed of the epidemic, whereas mass vaccination could in principle proceed as quickly as logistics allowed, thereby minimizing the duration of an outbreak.

Model-based risk assessment

A key benefit of using models to examine disease control options is their ability to explain and—with appropriate caveats—predict trends at a population level from interactions and processes at the individual level. Often the emergent dynamics at the population scale may be anything but obvious, owing to the many nonlinearities, complexities and feedbacks arising from the basic mechanisms at the individual level.

A corollary of the need to capture the mechanisms of transmission and control policies is that models have to be appropriately designed for the questions being addressed (Box 2). For instance, the feasibility of implementing different control policies is a key issue that will depend on the incidence of the disease. Models need to incorporate logistical constraints on policy implementation (for example, how many people can be vaccinated per day⁶) that are commensurate with that level of incidence. If the speed/risk of spread between communities is of interest, then models need to incorporate spatial structure. If contact tracing is to be assessed then some concept of social structure must be included.

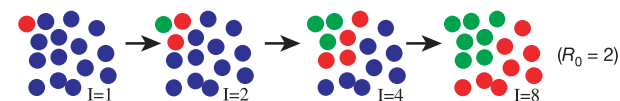
Table 2 compares model structures and assumptions of four recent studies^{5–8}. The diversity of model structures chosen and assumptions made regarding infection seeding, spread and control complicates direct comparison of the results of these studies. One issue for all the studies is the numbers of parameter values that are assumed, rather than estimated from data²¹. The most important

Box 1 Epidemics: basic theory

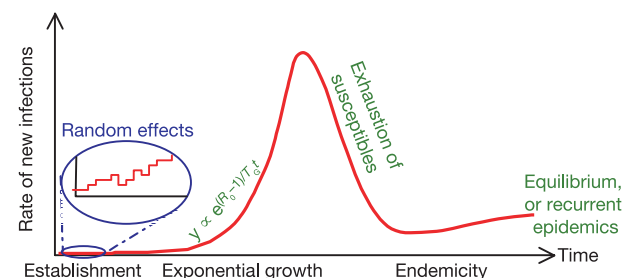
Individuals affected by an epidemic move through a number of infection states:



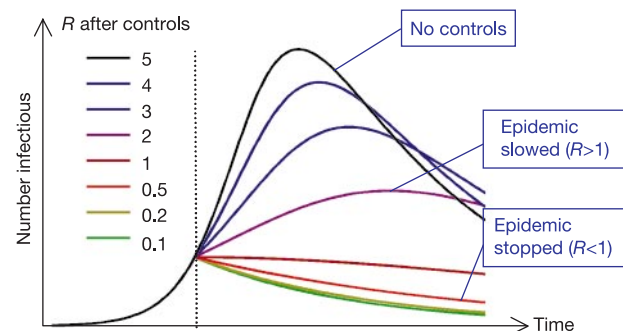
Epidemics are spread through contact (for example, person to person):



The 'chain reaction' nature of the process leads to exponential growth (once infected numbers are great enough to make random effects small), until the epidemic begins to run out of people to infect. This effect is summarized by the effective reproduction number R (ref. 11), which declines from its maximum, R_0 , as the susceptibles are depleted. Together, these processes give rise to the classic epidemic profile:



Controls either reduce susceptible numbers (such as vaccination) or limit transmission (for example, through movement controls). Both have the effect of reducing R and slowing the spread of an epidemic; reducing R below 1 means that the chains of transmission cannot be sustained and the epidemic dies out:



factors determining the conclusions drawn by two of the studies^{6,7} are the assumed infectiousness during prodrome relative to that during symptomatic disease, and the relatively low effectiveness of isolation of symptomatic cases assumed. New best estimates from a well-observed historical epidemic in Africa indicate prodromal transmission alone contributes 0.16 to R_0 (ref. 21), 2.4% of the total R_0 value of 6.9. Both of the studies assumed that prodromal R_0 is 2.5–3 out of a total R_0 of around 3 (refs 6, 7). Analyses using the new estimates show that isolation and ring vaccination—if logistically feasible—is nearly always optimal, and never markedly worse than mass vaccination at minimizing mortality²⁴.

The remaining two studies make simplifying assumptions that limit their usefulness. Reference 5 assumes purely exponential epidemic growth, meaning that little can be inferred about outbreaks of large size or duration. References 5 and 8 assume (rather than model) the impact of different control measures on transmission, meaning that their conclusions regarding control policy effectiveness are largely predetermined.

Data and uncertainty

The need to increase model sophistication and accuracy gives rise to an intrinsic tension: as model realism is increased the transparency associated with simple frameworks is often lost and the validation of model conclusions becomes harder. For models to be useful tools in policy planning, it is essential that they can be parameterized from

available data, and tested against past and current epidemic outbreaks, with proper consideration of changes in human populations with respect to immunity, mobility and patterns of social interaction.

Simple models have fewer parameters, which tends to make parameter estimation easier. Conversely, more complex models may have dozens of parameters describing the details of disease biology, host movement patterns and population structure. Unless all these parameters are robustly estimated or the effect of uncertainty in their values explored, there is a danger that incorrect assumptions will be made (and obscured by the complexity); this can make more detailed models no more reliable (and sometimes less so) than simpler frameworks. Achieving the correct balance between model complexity and validation is therefore key to informative modelling (Box 2 and Table 2). Validation in this context is how well the model matches observed epidemic behaviour at the level of detail relevant to the model's purposes. Ideally, such data should be independent of any epidemiological data used to estimate model parameters. When this ideal is not achievable (for example, when analysing an emerging epidemic of a novel disease^{25,26}), the use of rigorous statistical methods for assessing model goodness of fit is imperative.

The most useful data for parameterization and validation of models are detailed (individual case) reports from historical outbreaks^{2,3}. Such data provide the only reliable estimates of the reproduction number of smallpox (4–10, depending on the out-

Box 2

Modelling complexities

Greater disease realism

Extend the SIR model to incorporate the known within-host disease behaviour:



Duration: 12 days (±3) 3 days (±2) 9 days (±5) 17 days (±9)

Relative infectiousness in the prodromal and symptomatic periods is crucial in determining the optimal control strategy, with greater prodromal spread favouring mass vaccination.

Capturing social/spatial structure

Homogeneous mixing. Standard assumption of simple models—an individual has an equal chance of contacting anyone in the population. Contacts are independent (individuals who have previously made contact have no more or less chance of contacting again).

Age/social structure. Individuals have different probabilities of contact within specified population subgroups and between them. Subgroups might be defined by age, occupation (for example, hospital, school), socio-economic or health factors.

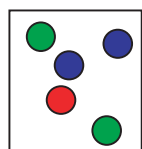
Network structure. Individuals form stable contact networks (for example, household, work colleagues, friends), the structure of which determines transmission dynamics. There can be rapid localized spread, followed by slow down as depletion of local susceptibles occurs due to correlations and clustering.

Patch structure. Populations are clumped, so towns/cities are natural units to study. Thus 'patch' (metapopulation) models are a valuable tool. There is more mixing within a patch than between patches, and patches can be out of sync in terms of epidemic progression.

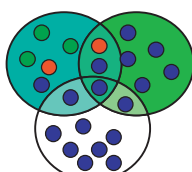
Individually based models. Stochastic simulation of contact patterns and disease progression at the level of individuals allows models to capture arbitrary levels of heterogeneity (including network and geographic) structure. However, such models are both computationally and data intensive if rigorous validation and parameter estimation is to be performed.

Stochastic or deterministic models?

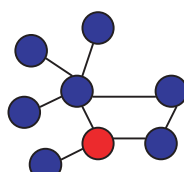
Deterministic (clockwork) models are rapid to simulate, relatively easy to parameterize, and hopefully capture the average epidemic behaviour. Stochastic models recognize the random nature of transmission events. As such they allow an assessment of the variability of the epidemic behaviour and are essential to deal with the low levels of infection near the start and end of an epidemic³⁰.



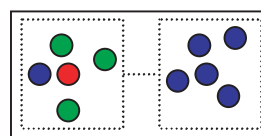
Homogenous mixing



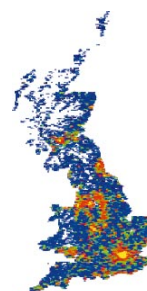
Age/social structure



Network structure



Patch structure



Individually based models

Table 1 Policy options for controlling a smallpox attack

Policy	Benefits	Drawbacks
Quarantine/isolation Quarantine and isolation of suspect and confirmed cases.	If isolation facilities are adequate, it is highly effective at reducing transmission from known cases	Isolation facilities necessary, or compliance with voluntary policy. Compulsory policies necessarily coercive. Requires rapid detection of cases.
Movement restrictions For example, quarantine of neighbourhoods or closure of schools, airports or other transport systems.	Potentially useful in containing a small outbreak where community transmission is occurring. Used recently to control SARS spread in Hong Kong and Singapore.	Vaccination certificates issued in the past to prevent potential spread but difficult to assess effectiveness. Costly and difficult to police, compromised by any illegal movements. Coercive.
'Ring' vaccination Contacts of suspect smallpox cases are traced and vaccinated when found. Can be coupled with policy of isolation of identified contacts.	Minimizes use of vaccine, and hence morbidity and mortality caused by adverse reactions to vaccination.	Contacts need to be found at an early stage of incubation for vaccine to be effective. Tracing needs to be highly effective to severely limit transmission.
Targeted vaccination For example, vaccination of whole population in affected neighbourhood or city.	Highly effective during eradication campaign at containing transmission localized to a single geographic area or subpopulation. Reduced vaccine-related mortality. Not dependent on contact tracing.	Effective when background levels of herd immunity high, but few systematic data on effectiveness in other contexts. Less sparing of vaccine use than ring vaccination. Risk of disease spreading beyond targeted area.
Mass vaccination Vaccination of whole population of a country experiencing or threatened by an outbreak.	Effective at stopping widespread dissemination of the virus across large areas and protecting individuals from infection. Not dependent on contact tracing.	Large numbers need to be vaccinated quickly. Might generate unnecessary vaccine-related morbidity and mortality. If policy implemented rapidly, screening for risk factors for adverse reactions might be suboptimal.
Prophylactic vaccination Vaccination before a smallpox release.	Useful for protecting essential 'first-responder' personnel. If used for entire population, very effective at stopping widespread dissemination of virus. Does not have to be implemented quickly. Not dependent on contact tracing.	If used to protect an entire population on an ongoing basis, policy has high, long-term cost, and a large number of probably unnecessary vaccine-associated adverse events would be expected for as long as policy is followed.

These policy options are unlikely to be applied in isolation of each other and will necessarily depend on availability of resources and levels of preparedness.

break^{20,21}), and information on how infectiousness varies at different disease stages^{2,21}. However, extrapolating from historical data to contemporary developed-world populations is problematic. It is unclear how much residual immunity remains today as a result of past vaccination programmes; those vaccinated 25–30 years ago are unlikely to possess complete immunity and a significant proportion may develop less severe forms of the disease^{2,3}, potentially changing the dynamics of transmission. Allowing for past immunity levels is therefore critical when estimating R_0 from historical data²⁰. Historically, most infections occurred in care-givers to symptomatic individuals, whether in households or hospitals³. It is unclear how

30 years of changes in household sizes, working patterns and mobility would affect transmission patterns today (see Box 2). Incorporating detailed data on demographics and human mobility into spatially explicit models offers one method by which such extrapolation can be made more reliable, but the scale of changes mean that much uncertainty will inevitably remain.

With such uncertainty, it is critical that risk assessment studies use modern statistical methods²⁷ to obtain the best possible parameter estimates from historical data, while allowing for changes in the last 30 years. However, historical data are less relevant for some key parameters—such as the likely scale of a bioterrorist attack, how

Table 2 Summary of recent smallpox modelling studies

Study	Key features	Comments
Meltzer <i>et al.</i> (ref. 5)	(1) Homogeneous mixing; (2) stochastic; (3) no social/spatial structure; (4) $R_0 = 1.5$ –3; (5) vaccination not directly modelled but in conjunction with quarantine is assumed to reduce R to below 1; (6) no depletion of susceptibles; (7) 100 people initially infected	(1) Estimates required control effort by correlating vaccine doses used against number of cases in historical outbreaks; (2) number of doses is not related to the number of cases, but to the level of susceptibility (and R_0); (3) model substantially overestimates cases because depletion of susceptibles is not modelled; (4) controls assumed to reduce R to 0.99, that is, programme effectiveness is model input
Kaplan <i>et al.</i> (ref. 6)	(1) Homogeneous mixing in large population (10 million); (2) deterministic; (3) no social/spatial structure; (4) $R_0 = 3$ for base case (range 1–20); (5) mass vaccination and ring vaccination compared; (6) considers public health logistical constraints; (7) only those 'asymptomatic' (that is in prodrome) are infectious—those symptomatic (with rash) are assumed to be isolated; (8) number initially infected = 1–100,000	(1) R_0 assumed derived from historical estimates, but level of prodromal transmission assumed much greater than new best estimates derived from historical data ²¹ ; (2) assumption that transmission occurs during the prodromal period biases results in favour of mass vaccination; (3) assumption of homogeneous mixing will also bias the results in favour of mass vaccination; (4) considers vaccine-related deaths only in contraindicated people, and not those in non-contraindicated individuals or other adverse events
Halloran <i>et al.</i> (ref. 7)	(1) Heterogeneous mixing (social structure included) in small population (2,000); (2) stochastic; (3) $R_0 = 3.2$; (4) mass vaccination and ring vaccination compared; (5) considers current residual herd immunity; (6) number initially infected = 1–10	(1) Large number of parameters assumed, particularly regarding mixing patterns, limited sensitivity analysis performed; (2) relatively large proportion (0.05–0.5%) of community initially infected; (3) model of small community of 2,000 individuals; (4) assumption that >75% of transmission occurs during prodrome biases results in favour of mass vaccination
Bozzette <i>et al.</i> (ref. 8)	(1) Homogeneous mixing; (2) stochastic; (3) R (no control) = 15, 3.4 and 1.8 in hospital, mixed and community outbreaks respectively; ($R_0 \geq R$ (no control)); (4) compares mass vaccination, ring vaccination and prophylactic vaccination of healthcare workers; (5) considers vaccine-related adverse events; (6) number initially infected = 2–100,000	(1) Assumes the effect of the control policies on R , based on review and 'judgement'; that is, effectiveness of policies is an input into the model rather than an output; the model therefore has little explanatory or predictive power; (2) large number of parameters assumed, limited sensitivity analysis performed; (3) assumes disease has no effect on the depletion of susceptibles; (4) assessment of threat of attack is subjective

rapidly the disease would be recognized, and the ability of public health authorities to respond. Furthermore, it is unclear how population behaviour would change in the face of an epidemic. Before recognition of the outbreak, would individuals in the latter stages of prodrome have more or fewer contacts, compared with their historical counterparts? Once smallpox is identified, will people voluntarily restrict their movements, or attempt to flee urban centres? These factors need to be explored with robust analysis^{28,29} of the sensitivity of model results and predicted optimal controls to parameter assumptions.

Particularly important is the assessment of the potential for catastrophic outcomes. A policy option may be optimal for the great majority of possible parameter scenarios, but fail to control disease spread in a few worst-case scenarios, which are still feasible given current data. An alternative policy option that is slightly less optimal for most scenarios but controls spread in the worst case might then be preferred.

For models to have a meaningful role in influencing policy decisions, it is therefore critical that not just 'most likely' or 'worst case' scenario modelling results are communicated but that a more detailed understanding of the sensitivity of predictions of outbreak size and policy optimality to model assumptions is conveyed, together with open acknowledgement of model or data weaknesses. Once an outbreak has begun, such questions may also be answered more precisely through the use of real-time modelling to refine parameters and better inform policy.

Conclusions

Given the many uncertainties outlined above, we argue that no model can be truly predictive in the context of smallpox outbreak planning, and no one control method can be identified *a priori* as best. Instead, modelling should aim to identify effective interventions for a variety of release scenarios that span the ranges of uncertainty in key parameters. By attempting to identify a single 'optimal' strategy, recently published studies are arguably attempting the impossible; this is reflected in their differing conclusions, which can largely be attributed to underlying differences in model structures and parameter assignments. A more useful goal for modelling is to identify a small set of control options that might be used in a range of scenarios, together with a set of trigger thresholds, which might determine when responses need to be escalated. The key is to match model structure and aims to determine the 'necessary' level of detail.

While recognizing the limitations of modelling for precise prediction, models represent a potentially powerful resource in the face of an actual outbreak. The 2001 foot-and-mouth disease epidemic in Britain^{14–17} highlighted the contribution that real-time statistical analysis and modelling could make in both predicting the future course of the outbreak and identifying the measures needed for its control. For smallpox, such a role might be even more critical—to identify which of the many parameter and release scenarios explored in preparatory planning is actually being realized (for example, in relation to changes in population behaviour caused by the outbreak), and thus which set of controls is likely to be optimal in containing the outbreak while minimizing casualties. The foot-and-mouth story highlighted the potential advantages to be gained from having model and data structures in place as much as possible before any epidemic. In this sense recent studies^{5–8} and the ensuing debate are to be welcomed.

Key to such a role is how much information would be available during the early stages of an epidemic to enable the analysis to be reliable; such data would also need to give enough early warning of a 'failing' intervention to allow intensification of controls sufficiently rapidly to prevent substantial excess deaths. Investigation of this issue now, using simulation and analytical studies, is critical. Real-time data capture and dissemination is essential, and infrastructure

needs to be developed in advance to facilitate this critical mechanism for implementing and monitoring interventions. Currently, it is unclear what level of detail would be necessary for models used in such a context—although some incorporation of the random effects that dominate epidemics in their early stages would clearly be desirable. If it were determined that the first few generations of cases in an outbreak would give insufficient information for real-time policy optimization, then a precautionary 'hit hard, hit early' policy might be warranted (coupled with a set of criteria for de-escalating the policy), despite the higher associated costs and adverse consequences of vaccinating more people than might strictly be necessary. Such considerations are of equal relevance to the emergence of new pathogens such as SARS^{25,26}, as they are to deliberately released agents. □

doi:10.1038/nature02007.

1. Lane, H. C., Montagne, J. L. & Fauci, A. S. Bioterrorism: a clear and present danger. *Nature Med.* **7**, 1271–1273 (2001).
2. Dixon, C. W. *Smallpox* (Churchill, London, 1962).
3. Fenner, F., Henderson, D. A., Arita, I., Jezek, Z. & Ladnyi, I. D. *Smallpox and its Eradication* (WHO, Geneva, 1988).
4. US sounds alarm over smallpox weapon threat. *Nature* **399**, 628 (1999).
5. Meltzer, M. I., Damon, I., LeDuc, J. W. & Millar, J. D. Modeling potential responses to smallpox as a bioterrorist weapon. *Emerg. Inf. Dis.* **7**, 959–969 (2001).
6. Kaplan, E. H., Craft, D. L. & Wein, L. M. Emergency response to a smallpox attack: The case for mass vaccination. *Proc. Natl Acad. Sci. USA* **99**, 10935–10940 (2002).
7. Halloran, M. E., Longini, I. M., Nizam, A. & Yang, Y. Containing bioterrorist smallpox. *Science* **298**, 1428–1432 (2002).
8. Bozzette, S. A. *et al.* A model for a smallpox-vaccination policy. *N. Engl. J. Med.* **348**, 416–425 (2003).
9. Lane, J. M. & Goldstein, J. Evaluation of 21st-century risks of smallpox vaccination and policy options. *Ann. Intern. Med.* **138**, 488–493 (2003).
10. Kemper, A. R., Davis, M. M. & Freed, G. L. Expected adverse events in a mass smallpox vaccination campaign. *Eff. Clin. Pract.* **5**, 84–90 (2002).
11. Anderson, R. M. & May, R. M. Directly transmitted infectious diseases: control by vaccination. *Science* **215**, 1053–1060 (1982).
12. Grenfell, B. T., Bjornstad, O. N. & Kappey, J. Travelling waves and spatial hierarchies in measles epidemics. *Nature* **414**, 716–723 (2001).
13. Anderson, R. M., May, R. M. & McLean, A. R. Possible demographic impact of AIDS in developing countries. *Nature* **332**, 228–234 (1988).
14. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. The foot-and-mouth epidemic in Great Britain: pattern of spread and impact of interventions. *Science* **292**, 1155–1160 (2001).
15. Keeling, M. J. *et al.* Dynamics of the 2001 UK foot and mouth epidemic: stochastic dispersal in a heterogeneous landscape. *Science* **294**, 813–817 (2001).
16. Keeling, M. J., Woolhouse, M. E., May, R. M., Davies, G. & Grenfell, B. T. Modelling vaccination strategies against foot-and-mouth disease. *Nature* **421**, 136–142 (2003).
17. Ferguson, N. M., Donnelly, C. A. & Anderson, R. M. Transmission intensity and impact of control policies on the foot and mouth epidemic in Great Britain. *Nature* **413**, 542–548 (2001).
18. Anderson, R. M. & May, R. M. *Infectious Diseases of Humans: Dynamics and Control* (Oxford Univ. Press, Oxford, 1991).
19. Wallinga, J., Levy-Bruhl, D., Gay, N. J. & Wachmann, C. H. Estimation of measles reproduction ratios and prospects for elimination of measles by vaccination in some Western European countries. *Epidemiol. Infect.* **127**, 281–295 (2001).
20. Gani, R. & Leach, S. Transmission potential of smallpox in contemporary populations. *Nature* **414**, 748–751 (2001).
21. Eichner, M. & Dietz, K. Transmission potential of smallpox: Estimates based on detailed data from an outbreak. *Am. J. Epidemiol.* **158**, 110–117 (2003).
22. Bailey, N. T. J. *The Mathematical Approach to Biology and Medicine* (Wiley, London, 1967).
23. Koopman, J. Controlling smallpox. *Science* **298**, 1342–1344 (2003).
24. Eichner, M. Case isolation and contact tracing can prevent the spread of smallpox. *Am. J. Epidemiol.* **158**, 118–128 (2003).
25. Lipsitch, M. *et al.* Transmission dynamics and control of severe acute respiratory syndrome. *Science* **300**, 1966–1970 (2003).
26. Riley, S. *et al.* Transmission dynamics of the etiological agent of SARS in Hong Kong: impact of public health interventions. *Science* **300**, 1961–1966 (2003).
27. O'Neill, P. D. A tutorial introduction to Bayesian inference for stochastic epidemic models using Markov chain Monte Carlo methods. *Math. Biosci.* **180**, 103–114 (2002).
28. Mollison, D. Simplifying simple epidemic models. *Nature* **310**, 224–225 (1984).
29. Bailey, N. T. J. & Duppenthaler, J. Sensitivity analysis in the modelling of infectious disease dynamics. *Math. Biosci.* **10**, 113–131 (1980).
30. Isham, V. & Medley, G. *Models for Infectious Human Disease: Their Structure in Relation to Data* (Cambridge Univ. Press, Cambridge, 1996).

Acknowledgements We thank the Royal Society (N.M.F. and M.J.K.), MRC (N.M.F.), Howard Hughes Medical Institute (N.M.F.), BBSRC (M.J.K. and B.T.G.), the Wellcome Trust (B.T.G. and R.M.A.) and the Department of Health (W.J.E., R.G. and S.L.) for funding. The views expressed are not necessarily those of the Department of Health, but we thank staff of the Department for discussions.

Correspondence and requests for materials should be addressed to N.M.F. (neil.ferguson@imperial.ac.uk).

Global analysis of protein localization in budding yeast

Won-Ki Huh^{1*}, James V. Falvo^{1*}, Luke C. Gerke¹, Adam S. Carroll¹, Russell W. Howson¹, Jonathan S. Weissman^{1,2} & Erin K. O'Shea¹

¹Howard Hughes Medical Institute, University of California–San Francisco, Department of Biochemistry and Biophysics, and ²Department of Cellular and Molecular Pharmacology, 600 16th Street, San Francisco, California 94143-2240, USA

* These authors contributed equally to this work

A fundamental goal of cell biology is to define the functions of proteins in the context of compartments that organize them in the cellular environment. Here we describe the construction and analysis of a collection of yeast strains expressing full-length, chromosomally tagged green fluorescent protein fusion proteins. We classify these proteins, representing 75% of the yeast proteome, into 22 distinct subcellular localization categories, and provide localization information for 70% of previously unlocalized proteins. Analysis of this high-resolution, high-coverage localization data set in the context of transcriptional, genetic, and protein–protein interaction data helps reveal the logic of transcriptional co-regulation, and provides a comprehensive view of interactions within and between organelles in eukaryotic cells.

Eukaryotic cells are organized into a complex network of membranes and compartments, which are specialized for various biological functions. Comprehensive knowledge of the location of proteins within these cellular microenvironments is critical for understanding their functions and interactions; this requires assaying the cell's full complement of proteins. The complete genome sequence of the budding yeast *Saccharomyces cerevisiae*¹ coupled with high-throughput experimental techniques has made systematic analyses of a eukaryotic proteome feasible. Recent studies have taken a genome-wide approach to analysing messenger RNA abundance and stability^{2,3}, biochemical activity^{4,5}, protein–protein interactions^{6–9}, transcriptional regulation¹⁰, gene disruption phenotypes^{11–14} and protein abundance¹⁵.

Previous large-scale analyses of protein localization in *S. cerevisiae* have depended on transposon-mediated random epitope tagging and plasmid-based overexpression of epitope-tagged proteins^{11,16}. However, epitope tagging of partial open reading frames (ORFs) can interrupt important localization signals, and overexpression of proteins may saturate intracellular transport mechanisms, leading to abnormal subcellular localization. To circumvent these potential problems, we generated a yeast strain collection expressing full-length proteins, tagged at the carboxy terminal end with green fluorescent protein (GFP), from their endogenous promoters by inserting the coding sequence of *Aequorea victoria* GFP (S65T)¹⁷ in-frame immediately preceding the stop codon of each ORF. With this strategy, wild-type levels and patterns of protein expression are minimally perturbed. Furthermore, because GFP fluorescence does not require external cofactors, GFP signal can be monitored in living cells without disrupting cellular integrity. We have analysed this strain collection using fluorescence microscopy to comprehensively characterize protein subcellular localization in a simple eukaryotic cell.

Construction and analysis of a GFP-tagged library

We systematically tagged each ORF in its chromosomal location through oligonucleotide-directed homologous recombination (Fig. 1a). For each of the 6,234 annotated ORFs¹⁸ a pair of oligonucleotides was generated that had homology to the desired chromosomal insertion site at the 5' end of each primer and homology to a vector containing the GFP tag at the 3' end. These primers were used to amplify the GFP tag and an auxotrophic marker from a plasmid template¹⁹, and the resulting polymerase chain reaction (PCR) products were transformed into a haploid

yeast strain. Transformants were assayed by genomic PCR with one primer specific for the GFP tag and a second specific for each ORF, to determine whether the cassette had integrated at the appropriate locus. A total of 6,029 strains with chromosomally GFP-tagged ORFs were grown to mid-logarithmic phase in synthetic medium and analysed by fluorescence microscopy; 4,156 of these showed GFP signals above background levels (Table 1).

Micrographs of each GFP-tagged strain (Fig. 1b; see also Supplementary Fig. S1), lacking ORF identifiers, were independently evaluated by two scorers and initially classified into one or more of 12 subcellular localization categories (Table 2). We then refined these categories by performing a series of co-localization experiments. Haploid reference strains expressing monomeric red fluorescent protein (mRFP)²⁰ fusions to proteins whose localization had been characterized previously (Table 2) were mated to approximately 700 GFP strains that were not assigned definitive localizations by GFP microscopy alone, and the resulting diploid cells were analysed by fluorescence microscopy (Fig. 1c). On the basis of this analysis, proteins were assigned to an additional 11 localization categories (Table 2). All information was captured into a database (<http://yeastgfp.ucsf.edu>).

Subcellular localization of yeast proteins

The 4,156 proteins for which we defined subcellular localizations in the GFP library represent 75% of the yeast proteome^{15,21,22}. Our results provide localization data for about 70% of previously unlocalized yeast proteins, constituting about 30% of the proteome (Fig. 2a). Over 90% of the proteins visible in the GFP collection were also detected by western blot analysis of a collection of TAP (tandem affinity purification)-tagged strains¹⁵, suggesting that the false-positive rate in this study is extremely low.

The distribution of protein subcellular localization reveals that, as expected, many proteins are found in the nucleus or cytoplasm, whereas 1,839 proteins, 44% of the total observed, localize to other specific subcellular regions (Fig. 2b). Notably, over 40% of the proteins that we assigned to the cytoplasm, late Golgi/clathrin and lipid particle represent new localization assignments. There are limitations to the subcellular localizations in yeast discernible by fluorescence microscopy; for example, we cannot distinguish kinetochore versus spindle pole body, or membrane versus lumen for mitochondria or the endoplasmic reticulum. However, use of the GFP tag and co-localization with RFP-tagged reference proteins allowed us to resolve many related subcellular compartments with

confidence. For example, the nucleus, nuclear periphery and the endoplasmic reticulum are distinct (Fig. 1b, top row), as are the vacuole and vacuolar membrane, and multiple compartments of the secretory pathway. This level of precision greatly facilitates our assignment of protein localization as well as integration with other genome-wide data sets.

Previously published localization data from the *Saccharomyces* Genome Database (SGD)¹⁸, including data from earlier large-scale studies^{11,16}, were available for a total of 2,526 proteins visible in the GFP library—we found that there was 80% agreement between our data and those of the SGD. We also found that our localization assignments generally agree with those of the pioneering studies of

the Snyder laboratory^{11,16}. However, for those assignments that differ, our results show closer agreement with the SGD (Supplementary Fig. S2). Direct comparison between our data and the results of a mass spectrometric analysis of the nuclear pore complex²³ (NPC) revealed that, of 29 identified NPC components, 25 were visible in our study: 23 proteins (92%) were localized to the nuclear periphery and one each was localized to the nucleus/cytoplasm and endoplasmic reticulum. Furthermore, of 16 spindle-pole-body components identified by mass spectrometry²⁴, all 14 of the proteins visible in this study were localized to the spindle pole. We found an additional 20 proteins localized to the nuclear periphery and 14 to the spindle pole; of these, 11 had not been detected previously in the nuclear periphery and 7 had not been detected in the spindle pole (Supplementary Table S1). The strong correlation between the data we obtained by fluorescence microscopy and localization data obtained by other methods supports the reliability of this study in defining new protein localizations.

A potential source of discrepancy between our data and those from other studies is that the C-terminal fusion of the GFP protein (approximately 27 kDa) may cause mislocalization through steric hindrance or interruption of critical C-terminal localization/retention sequences (Supplementary Table S2). For example, the small GTP-binding protein Ras2 was localized to the nucleus and the cytoplasm in this study, but it is known to be localized to the plasma membrane due to modification of its C terminus with palmitoyl and farnesyl groups²⁵. Proteins localized to the cell wall²⁶ and subsets of proteins localized to the peroxisome²⁷ and endoplasmic reticulum²⁸ also contain C-terminal targeting signals, and these were often mislocalized in this study.

Organellar proteomics of the nucleolus

The identification of subsets of proteins in various organelles is an initial step towards the understanding of biological processes at the cellular level. 'Organellar proteomics' studies²⁹ would benefit especially from the comprehensive localization data for yeast proteins provided by this study. For example, we detected 164 proteins in the nucleolus in this study; 82 of these overlap the 127 nucleolar proteins catalogued in the SGD, but another 82 are newly defined (Fig. 2c). Of the remaining 45 nucleolar proteins from the SGD, 28 were not visualized in our study, whereas the others were localized to the nucleus (7 proteins), nucleus/cytoplasm (7 proteins), nuclear periphery (2 proteins) and cytoplasm (1 protein). These proteins may occupy the nucleolus in a transient fashion, at levels not detectable by our methods, or under conditions distinct from those of our study—mislocalization may also result from the GFP tag. A number of the nucleolar proteins found in this study are involved in ribosomal RNA transcription and processing and in ribosome biogenesis, in accordance with the classical role of the nucleolus; for some of these proteins, we provide the first direct demonstration that they reside or are enriched in the nucleolus (Fig. 2d). Given that some nucleolar proteins are involved in cell cycle control and gene regulation^{30–33}, it will be very interesting to investigate the functional roles of nucleolar proteins newly defined in this study.

It has been reported that essential proteins and orthologues are

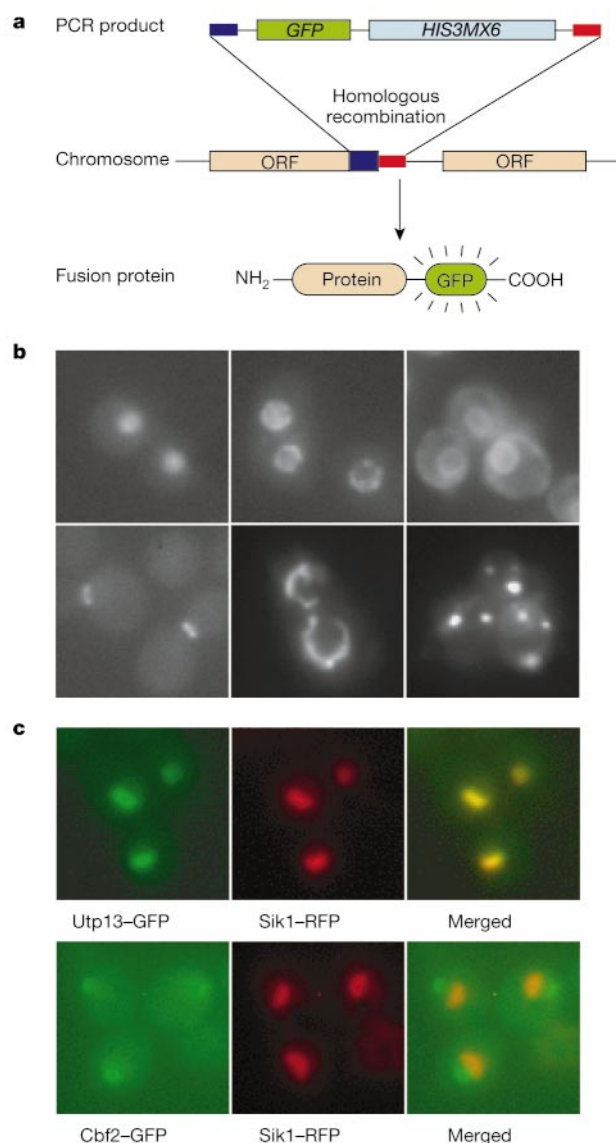


Figure 1 Microscopic analysis of yeast strains expressing GFP-tagged proteins. Data and images are accessible at <http://yeastgfp.ucsf.edu>. **a**, Strategy for library construction. PCR products containing the GFP tag and a selectable marker gene were inserted at the C terminus of each ORF through homologous recombination, yielding a C-terminally GFP-tagged protein. **b**, Representative GFP images of Rox3-GFP (nucleus; top left), Nic96-GFP (nuclear periphery; top middle), Pho86-GFP (endoplasmic reticulum; top right), Hof1-GFP (bud neck; bottom left), Ilv6-GFP (mitochondrion; bottom middle) and Erg6-GFP (lipid particle; bottom right). **c**, Representative co-localization experiment. A Utp13-GFP or Cbf2-GFP yeast strain was mated with a strain containing Sik1-RFP as a nucleolar marker, then fluorescence images for GFP (left) and RFP (middle) were taken and merged (right).

ORF category	Number of ORFs	Success rate (%)
ORFs processed for PCR of the GFP cassette	6,234* (1,100)	100 (100)
ORFs with successful transformation	6,151 (1,018)	99 (93)
ORFs with positive homologous recombination confirmed by genomic PCR	6,029 (953)	97 (87)
ORFs with positive GFP signal	4,156 (827)	67 (75)

Values for essential ORFs are indicated in parentheses.
*ORFs annotated in SGD, 17 April 2001.

enriched in related protein complexes isolated from yeast and humans⁸. Of the proteins localized to the nucleolus in this study, 99 proteins (60%) are known to be essential, substantially more than the 20% required for viability in the proteome as a whole^{12,14}. Recently, mass spectrometric analysis of the human nucleolus identified 271 proteins, 166 of which have homologues in yeast³⁴; 52 of these proteins are classified as nucleolar in this study (Supplementary Table S3). Of the 112 proteins remaining from the 164 proteins that we have detected in the yeast nucleolus, 73 have human homologues and 33 of these are localized to the nucleolus or have biological functions related to transcription and processing of rRNAs and ribosome biogenesis (Supplementary Table S4) according to the Human Proteome Survey Database³⁵. Given the enrichment of essential proteins in the yeast nucleolus and the enrichment of essential proteins and orthologues in related protein complexes from yeast and humans⁸, we expect that many of the remaining human homologues of yeast proteins detected in the nucleolus in this study will also be nucleolar proteins.

Protein localization and mRNA co-expression

Many genome-wide analyses have demonstrated that mRNA transcript expression patterns are similar for groups of functionally related genes^{2,36–38}; mRNA abundance is also similar within certain cellular compartments³⁹. However, transcriptional co-regulation has not been directly compared to subcellular protein localization on a proteome-wide scale. To assess the extent of this correlation, we made use of a study that identified 33 transcriptional ‘modules’ of genes with marked co-regulation based on analysis of over 1,000 microarray data sets reflecting the results of different mutant strain backgrounds or environmental perturbations^{38,40}. For each module, the fraction of proteins with a given subcellular localization was calculated and divided by that fraction in the whole proteome to generate fold enrichment in each subcellular localization category (Fig. 3a). We obtained statistically significant enrichments (one-sided binomial test with $P < 0.05$) for 19 of the 22 most highly expressed modules, indicating that co-localization is strongly correlated with transcriptional co-expression and, by extension, with biological function.

The combination of protein localization and transcriptional co-expression can be used to corroborate or predict the function of unnamed ORFs in a specific module. For example, YGL068W and YNL122C, both of which belong to the mitochondrial ribosomal protein transcriptional module, localize to the mitochondrion in our study, as do 13 other members of this module, strongly supporting the function predicted by the module (Fig. 3b). Indeed, the sequence of YGL068W shows 49% similarity to that of the human mitochondrial ribosomal protein L12 (ref. 41).

Localization and co-regulation data can also be used to gain

insight into biological function when proteins in a given transcription module are enriched in more than one localization category. This allows us to subdivide sets of co-expressed proteins, providing a level of information that cannot be gleaned solely from their classification in the same module based on their expression profiles. For example, proteins in the G1 module (representing processes coordinated at the G1/S transition) localize to three basic categories: nucleus, bud/bud neck and spindle pole (Fig. 3c). The basic functions of proteins in the G1 module, where known, can be divided by localization; proteins localized to the bud/bud neck are involved in bud formation, whereas nuclear proteins from this module are involved mainly in chromosome cohesion, transcription, and DNA replication, repair and recombination. Thus, given

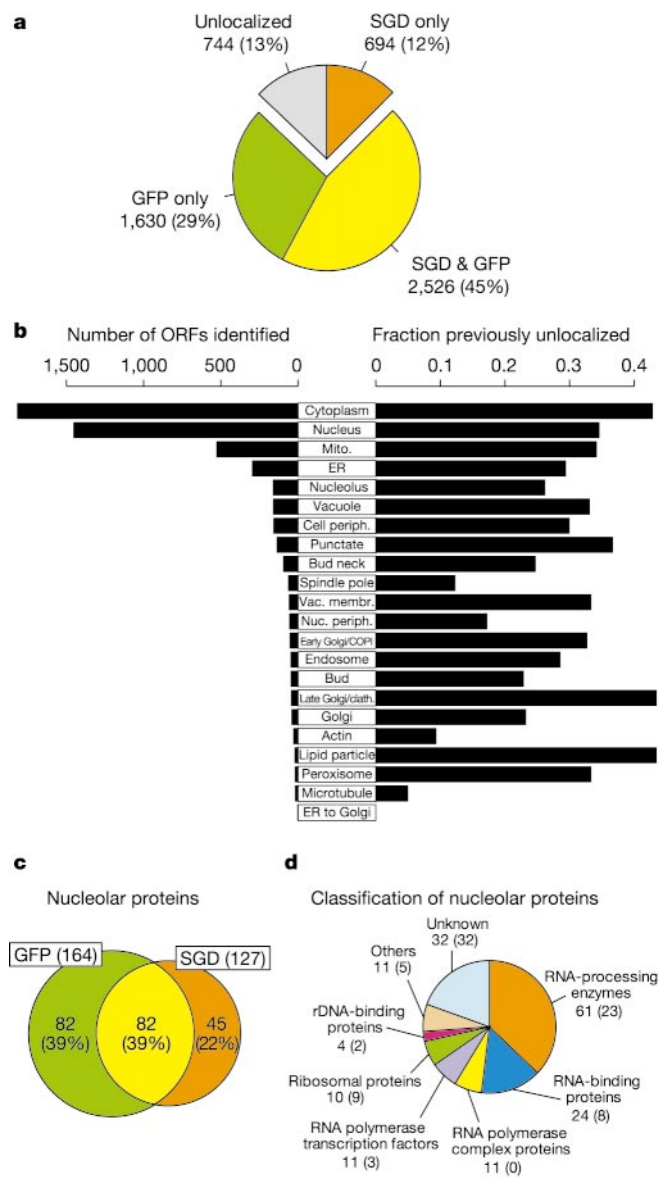


Figure 2 Subcellular localization of yeast proteins. **a**, Contribution of this study to expanding localization data for the yeast proteome. Six hundred and forty ORFs thought to be spurious¹⁵ were excluded from the pie chart. **b**, Distribution of the subcellular localizations for proteins visualized in this study. Also shown is the fraction of previously unlocalized proteins in each category. ER, endoplasmic reticulum. **c**, Comparison of the nucleolar proteins identified from this study with those from SGD. **d**, Functional classification of 164 nucleolar proteins defined in this study¹⁸. The numbers of ORFs with previously unknown subcellular localizations are shown in parentheses.

Table 2 Subcellular localization categories used in this study

GFP localization	GFP and RFP co-localization
Cell periphery	Cytoskeleton
Bud	Actin cytoskeleton
Bud neck	Spindle pole
Cytoskeleton	Nucleolus
Microtubule	Nuclear periphery
Cytoplasm	Golgi apparatus
Nucleus*	Transport vesicle
Mitochondrion*	Early Golgi/COPI
Endoplasmic reticulum	Late Golgi/clathrin
Vacuole	Endoplasmic reticulum to Golgi
Vacuolar membrane	Endosome
Punctate	Peroxisome
Ambiguous	Lipid particle

Co-localization experiments used the following RFP-tagged markers: Sac6 (actin cytoskeleton), Spc42 (spindle pole), Slt1 (nucleolus), Nic96 (nuclear periphery), Anp1 (Golgi apparatus), Cop1 (early Golgi/COPI), Chc1 (late Golgi/clathrin), Sec13 (endoplasmic reticulum to Golgi), Snf7 (endosome), Pex3 (peroxisome) and Erg6 (lipid particle).

*Initially assigned by DAPI staining; some punctate proteins were subsequently confirmed as mitochondrial using MitoTracker red.

that the G1 module proteins Hif1, Hsn1, YGR151C, YKR077W and YMR144W are localized to the nucleus, it is likely that they share the functions of nuclear proteins from this module.

Comparison with genetic and physical interactions

Recent genome-wide studies have sought to enumerate all protein–protein interactions that occur in *S. cerevisiae*^{6–9}. Despite the large scale of these efforts, the agreement between studies⁴² suggests that total coverage is poor and false-positive rates remain high. To interact physically proteins must exist in close proximity, at least transiently, suggesting that co-localization may be an effective means for evaluating hypothetical interactions. To assess the relationship between co-localization and interaction, we chose as a reference set the sum of all genetic and protein–protein interactions reported in the GRID database⁴³. Although this set is certain to contain a considerable fraction of false-positive interactions, it was chosen to minimize systematic bias in individual screens that inevitably results from alternative interaction detection methods. We determined the subcellular localizations of each interacting protein pair from this reference set and the fraction of the total number of interactions occurring for each localization pair. A set of randomized protein pairs was also generated from the whole proteome, and localization pair statistics were collected on this set in the same way. We calculated the fold enrichment observed for each localization pair in our reference data set as compared with the randomized data set to generate an interaction matrix (Fig. 4a).

This analysis supports and extends interaction data from other studies. As expected, interactions are strongly enriched between proteins that co-localize (one-sided binomial test with $P < 0.001$), but the degree of enrichment varies widely by compartment. For example, interactions between cytoplasmic proteins are 1.3-fold enriched above chance, whereas interactions between microtubule

proteins are 56-fold enriched above chance, implying that co-localization of two putative interacting proteins to the microtubule cytoskeleton provides better evidence of physical and functional interaction than the simple fact that they do co-localize.

Of particular interest is that enrichment in interactions was observed between distinct localization categories in our study, shown as red off-diagonal circles in the matrix (Fig. 4a). Such off-diagonal circles are indicative of functional relationships between subcellular localizations: they are neither the result of systematic errors from individual interaction data sets (Supplementary Fig. S3) nor of proteins being assigned multiple localizations (Supplementary Fig. S4). An extensive network of interactions occurs between proteins localized to the actin cytoskeleton and those localized to the bud neck (Fig. 4b). Notably, some proteins not previously known to localize to these regions have known or predicted functions consistent with their assigned localizations. The Rho-GTPase activator protein Bem2, for example, has been shown through genetic studies to be involved in bud growth, establishment of cell polarity, and organization and biogenesis of the actin and microtubule cytoskeleton and of the cell wall⁴⁴. These functions are consistent both with localization to the bud neck and functional interaction with the actin cytoskeleton. Similarly, Chs7 is involved in cell wall chitin biosynthesis⁴⁵, consistent with the role of certain bud neck proteins⁴⁶, and Akl1 has a predicted role in the organization of the actin cytoskeleton⁴⁷.

The biological importance of statistically significant interactions between localization categories in the GFP library is fully revealed when these interactions are considered in the context of a eukaryotic cell (Fig. 4c). A network of interactions connects subcellular regions that are functionally and physically related. Strongly interconnected localizations can reflect dynamic interchange of proteins between compartments; for example, compartments of the secretory path-

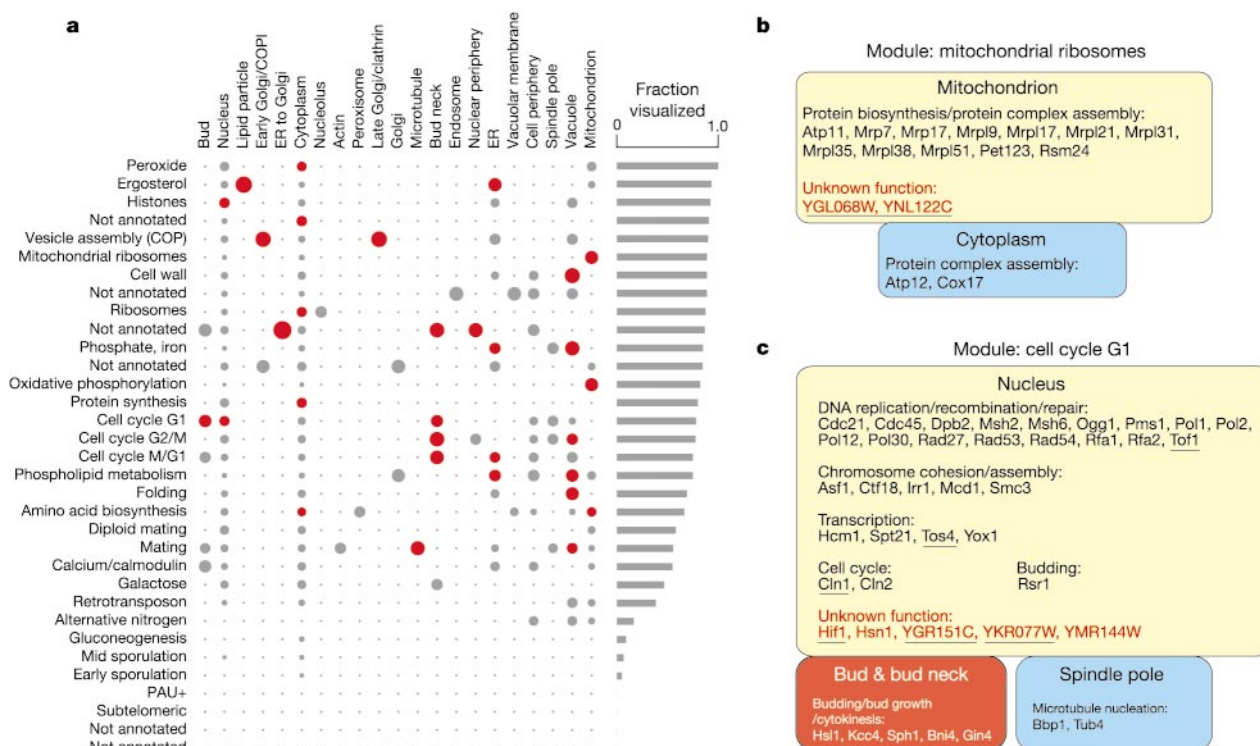


Figure 3 Correlation between transcriptional co-regulation and subcellular localization. **a**, The enrichment of subcellular localizations (top) exhibited by proteins belonging to each transcription module³⁸ (left) designated by biological function where applicable. Log(enrichment) is proportional to the radius of the circles shown; red circles indicate >95% confidence (one-sided binomial test) that enrichment >1. The fraction of proteins

visualized by GFP within each module is indicated by the bar graph (right). **b, c**, Diagrams of the proteins of the mitochondrial ribosome (**b**) and G1 (**c**) transcriptional modules divided into principal localizations observed in the GFP library, and further grouped by previously defined biological functions¹⁸. Proteins newly localized in this study are underlined, and proteins with unknown biological function are noted in red.

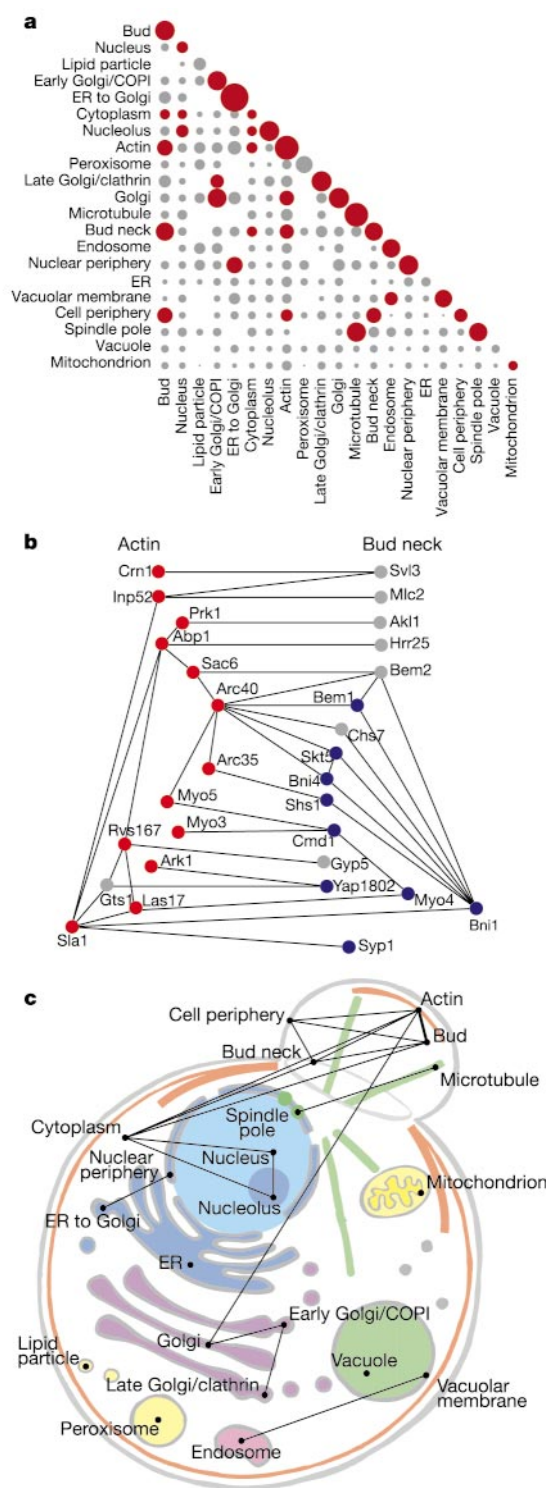


Figure 4 Relationship between genetic and physical interactions and subcellular localization. **a**, Localization pairs were extracted from the localization of interacting partners. Interaction between specific compartments was compared to a randomized interaction set, showing that some compartments interact preferentially with others. Log(enrichment) is proportional to the radii of the circles; red circles have >99.9% confidence (one-sided binomial test) that enrichment >1. ER, endoplasmic reticulum. **b**, Diagram⁴⁸ of genetic and protein–protein interactions represented by the off-diagonal enrichment between the actin cytoskeleton (left, red circles) and bud neck (right, blue circles). Grey circles denote proteins not localized previously to either category. **c**, Diagram of the off-diagonal, statistically significant interacting compartments in **a**.

way (Golgi, early Golgi/COPI (coat protein I) and late Golgi/clathrin). Intercompartmental interactions can also reflect close proximity and extensive physical association between localization categories, as is the case for the bud neck, the bud and the actin cytoskeleton. The interaction matrix provides an overview of communication between subcellular compartments as well as a template for evaluating the validity of protein–protein interactions from large-scale experimental or theoretical data sets.

Discussion

By creating a GFP-tagged yeast strain collection and database that covers three-quarters of the proteome and over two-thirds of previously unlocalized proteins, we have provided an experimental and informational resource to the scientific community. Although we have presented an analysis of the yeast proteome in a nominal resting state, the GFP library serves as a starting point for understanding the complex state of flux in the eukaryotic proteome that underlies the survival and development of an organism. The library provides a tool for analysing the global dynamics of the proteome in response to specific external stimuli or growth conditions over a selected period of time. Similarly, the library can be used in combination with high-throughput strain construction techniques¹³ to assay the effects of deletion or mutation of a protein of interest on global protein localization. Complex regulatory networks responsible for targeting proteins to specific cellular compartments can thus be systematically dissected.

We have shown that the combination of high-resolution, high-accuracy, proteome-wide localization information with data from other proteomics-scale studies provides an independent dimension of information that reveals patterns not visible within a single data set. The localization data from the GFP library can confirm and extend predictions based on trends within a single data set; if proteins grouped together in a given data set have a common localization, the prediction of common function is strengthened. This is particularly useful in the case of proteins for which little functional data exists. The localization data also make it possible to subdivide groups of proteins related by genome-wide trends in other data sets, indicating that one group may be composed of subsets of proteins with even more specific, separate biological roles. A comparative proteomics approach promises to reveal important features of basic cellular processes, improving our understanding of *S. cerevisiae* and of the proteins and pathways conserved among eukaryotes. □

Methods

Construction of GFP-tagged yeast strains

To construct a chromosomally GFP-tagged library, 6,234 pairs of gene-specific oligonucleotide primers were synthesized, each of which had been designed to share complementary sequences to the GFP tag-marker cassette at the 3' end and contain 40 base pairs (bp) of homology with a specific gene of interest to allow in-frame fusion of the GFP tag at the C-terminal coding region of the gene. Gene-specific cassettes containing a C-terminally positioned GFP tag were then generated by PCR using as a template pFA6a–GFP(S65T)–His3MX, which contains the *Schizosaccharomyces pombe* *his5*⁺ gene and permits selection of transformed strains in histidine-free media¹⁹. The haploid parent yeast strain (ATCC 201388: *MATa his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) was transformed with the PCR products, and strains were selected in SD medium (synthetic medium plus dextrose, Difco) lacking histidine. Insertion of the cassette by homologous recombination was verified by genomic PCR of samples from individual colonies with a primer internal to the GFP tag and a separate set of ORF-specific primers designed to produce a product of approximately 500 bp. Strains representing 6,029 ORFs were successfully tagged with GFP (Table 1), and independent strains from two to six selected colonies from each ORF were analysed by fluorescence microscopy.

Microscopic imaging of GFP-tagged strains

Aliquots of strains grown to mid-logarithmic phase in SD medium lacking histidine were analysed in 96-well glass-bottom microscope slides (BD Falcon) pre-treated with concanavalin A (50 μg ml^{−1}) to ensure cell adhesion. Cells were incubated in SD medium containing 1 μg ml^{−1} 4',6-diamidino-2-phenylindole (DAPI) as a marker for the nucleus and mitochondria, and analysed by multiple wavelength fluorescence and visible light microscopy with a digital imaging-capable Nikon TE200/300 inverted microscope using an oil-immersed objective at ×100 magnification. Using a script in MetaMorph version

4.6r8 imaging software (Universal Imaging Corporation), fluorescence microscopy images for GFP, DAPI and Nomarski/DIC (differential interference contrast) images were taken in rapid succession, and the stage was automatically advanced between wells on the 96-well slide.

Localization category refinement by co-localization

Subcellular localizations that could not be assigned readily by GFP fluorescence alone—typically classified as punctate or non-uniform nuclear—were resolved by mating the GFP-tagged strains to strains expressing reference proteins (Table 2) fused to mRFP²⁰. The coding sequence for mRFP was amplified by PCR from its parent vector (pRSET-mRFP1)²⁰ and inserted into pFA6a-KanMX6, which carries the *Escherichia coli* kanamycin-resistance gene¹⁹, to create the plasmid pFA6a-mRFP-KanMX6. This vector was then used to generate gene-specific cassettes to yield reference strains expressing C-terminally mRFP-tagged proteins. The haploid parent strain (ATCC 201389: *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*) was transformed, selected in the presence of G418 sulphate (200 µg ml⁻¹), and analysed for positive RFP signal by fluorescence microscopy as described above. The GFP-tagged strains were then mated with at least one of the mRFP-tagged reference strains in SD medium lacking lysine and methionine, and the resulting diploid strains were analysed by microscopy to generate GFP, RFP, DIC and GFP-RFP merged images. Haploid strains exhibiting potential non-uniform mitochondrial GFP patterns were subjected to the same microscopic analysis using the mitochondrion-specific dye MitoTracker red CMXRos (Molecular Probes).

Database features

We have designed a publicly available web-based user interface to the localization database at <http://yeastgfp.ucsf.edu>. At this site, users can perform searches using a number of criteria, including ORF name, gene name, subcellular localization, cell cycle, cell morphology, cell-cell brightness variability and subcellular signal heterogeneity. Searches retrieve full-sized, lossless compressed images that were used to assign localizations in this study; specific cells used to justify localizations are indicated in the images.

Comparison with other data sets

The distribution of subcellular localizations exhibited by the test ORF set (groups of transcriptionally co-regulated genes^{38,40}) was assessed in comparison to a reference set (the localization distribution seen in all ORFs characterized in this study). The identity of genes in the modules can be found at <http://barkai-serv.weizmann.ac.il/modules/page/details.html> using a threshold cutoff of 4.0. The frequency with which each subcellular localization is observed in the test and reference set was calculated; the ratio of these frequencies is reported as the enrichment. Individual binomial tests were performed for each subcellular localization to accept or reject the null hypothesis that the measured enrichment occurred due to chance. A one-tailed *P*-value <0.05 is taken to be statistically significant and is indicated by red circles in Fig. 3a. Distribution of subcellular localization of interacting partners was assessed by comparison to that which would occur by random association of ORFs, giving an enrichment of interactions between localizations. Individual binomial tests confirm that enrichment for certain localization pairs, indicated in Fig. 4a by red circles, is not the product of sampling error (*P* < 0.001).

Received 28 July; accepted 1 September 2003; doi:10.1038/nature02026.

- Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
- Velculescu, V. E. *et al.* Characterization of the yeast transcriptome. *Cell* **88**, 243–251 (1997).
- Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci. USA* **99**, 5860–5865 (2002).
- Martzen, M. R. *et al.* A biochemical genomics approach for identifying genes by the activity of their products. *Science* **286**, 1153–1155 (1999).
- Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293**, 2101–2105 (2001).
- Uetz, P. *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**, 623–627 (2000).
- Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98**, 4569–4574 (2001).
- Gavin, A.-C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
- Ho, Y. *et al.* Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**, 180–183 (2002).
- Lee, T. I. *et al.* Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804 (2002).
- Ross-Macdonald, P. *et al.* Large-scale analysis of the yeast genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
- Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
- Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294**, 2364–2368 (2001).
- Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418**, 387–391 (2002).
- Ghaemmaghami, S. *et al.* Global analysis of protein expression in yeast. *Nature* **425**, 737–741 (2003).
- Kumar, A. *et al.* Subcellular localization of the yeast proteome. *Genes Dev.* **16**, 707–719 (2002).
- Tsien, R. Y. The green fluorescent protein. *Annu. Rev. Biochem.* **67**, 509–544 (1998).
- Dolinski, K. *et al.* *Saccharomyces Genome Database* (<http://www-genome.stanford.edu/Saccharomyces/>) (2003).
- Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
- Campbell, R. E. *et al.* A monomeric red fluorescent protein. *Proc. Natl Acad. Sci. USA* **99**, 7877–7882 (2002).

- Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
- Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
- Rout, M. P. *et al.* The yeast nuclear pore complex: composition, architecture, and transport mechanism. *J. Cell Biol.* **148**, 635–651 (2000).
- Wigge, P. A. *et al.* Analysis of the *Saccharomyces* spindle pole by matrix-assisted laser desorption/ionization (MALDI) mass spectrometry. *J. Cell Biol.* **141**, 967–977 (1998).
- Bhattacharya, S., Chen, L., Broach, J. R. & Powers, S. Ras membrane targeting is essential for glucose signaling but not for viability in yeast. *Proc. Natl Acad. Sci. USA* **92**, 2984–2988 (1995).
- van Berkel, M. A., Caro, L. H., Montijn, R. C. & Klis, F. M. Glucosylation of chimeric proteins in the cell wall of *Saccharomyces cerevisiae*. *FEBS Lett.* **349**, 135–138 (1994).
- Gould, S. J. *et al.* Peroxisomal protein import is conserved between yeast, plants, insects and mammals. *EMBO J.* **9**, 85–90 (1990).
- Pelham, H. R., Hardwick, K. G. & Lewis, M. J. Sorting of soluble ER proteins in yeast. *EMBO J.* **7**, 1757–1762 (1988).
- Taylor, S. W., Fahy, E. & Ghosh, S. S. Global organellar proteomics. *Trends Biotechnol.* **21**, 82–88 (2003).
- Shou, W. *et al.* Exit from mitosis is triggered by Tem1-dependent release of the protein phosphatase Cdc14 from nucleolar RENT complex. *Cell* **97**, 233–244 (1999).
- Visintin, R., Hwang, E. S. & Amon, A. Cfi1 prevents premature exit from mitosis by anchoring Cdc14 phosphatase in the nucleolus. *Nature* **398**, 818–823 (1999).
- San-Segundo, P. A. & Roeder, G. S. Pch2 links chromatin silencing to meiotic checkpoint control. *Cell* **97**, 313–324 (1999).
- Rabitsch, K. P. *et al.* Kinetochore recruitment of two nucleolar proteins is required for homolog segregation in meiosis I. *Dev. Cell* **4**, 535–548 (2003).
- Andersen, J. S. *et al.* Directed proteomic analysis of the human nucleolus. *Curr. Biol.* **12**, 1–11 (2002).
- Hodges, P. E. *et al.* Annotating the human proteome: the Human Proteome Survey Database (HumanPSD) and an in-depth target database for G protein-coupled receptors (GPCR-PD) from Incyte Genomics. *Nucleic Acids Res.* **30**, 137–141 (2002).
- DeRisi, J. L., Iyer, V. R. & Brown, P. O. Exploring the metabolic and genetic control of gene expression on a genomic scale. *Science* **278**, 680–686 (1997).
- Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
- Ihmels, J. *et al.* Revealing modular organization in the yeast transcriptional network. *Nature Genet.* **31**, 370–377 (2002).
- Drawid, A., Jansen, R. & Gerstein, M. Genome-wide analysis relating expression level with protein subcellular localization. *Trends Genet.* **16**, 426–430 (2000).
- Bergmann, S., Ihmels, J. & Barkai, N. Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E* **67**, 031902 (2003).
- Koc, E. C. *et al.* The large subunit of the mammalian mitochondrial ribosome. *J. Biol. Chem.* **276**, 43958–43969 (2001).
- von Mering, C. *et al.* Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**, 399–403 (2002).
- Breitkreutz, B.-J., Stark, C. & Tyers, M. The GRID: the General Repository for Interaction Datasets. *Genome Biol.* **4**, R23 (2003).
- Madden, K. & Snyder, M. Cell polarity and morphogenesis in budding yeast. *Annu. Rev. Microbiol.* **52**, 687–744 (1998).
- Trilla, J. A., Duran, A. & Roncero, C. Chs7p, a new protein involved in the control of protein export from the endoplasmic reticulum that is specifically engaged in the regulation of chitin synthesis in *Saccharomyces cerevisiae*. *J. Cell Biol.* **145**, 1153–1163 (1999).
- DeMarini, D. J. *et al.* A septin-based hierarchy of proteins required for localized deposition of chitin in the *Saccharomyces cerevisiae* cell wall. *J. Cell Biol.* **139**, 75–93 (1997).
- Cope, M. J., Yang, S., Shang, C. & Drubin, D. G. Novel protein kinases Ark1p and Prk1p associate with and regulate the cortical actin cytoskeleton in budding yeast. *J. Cell Biol.* **144**, 1203–1218 (1999).
- Breitkreutz, B.-J., Stark, C. & Tyers, M. Osprey: a network visualization system. *Genome Biol.* **4**, R22 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank F. Lam for designing the original relational data model and for assistance with database organization; M. Springer for sharing microscopy expertise; J. Newman for assistance with FACS analysis; N. Barkai, S. Ghaemmaghami and K. Bower for discussions and sharing results; R. Tsien for providing the pRSET-mRFP1 plasmid; M. Levin for assistance with preparation of RFP-tagged strains; R. Marion for contributing GFP-tagged strains of CAK1, PHO4, RTS2, SET1 and STP4; A. DePace for preparing the illustration in Fig. 4c; D. Ahern, F. Sanchez, A. Belle and M. Liku for primer synthesis and other technical assistance; and S. Emr and members of the O'Shea and Weissman laboratories for critical discussion of the work. This work was supported by the Howard Hughes Medical Institute and the David and Lucile Packard Foundation (J.S.W. and E.K.O.). J.V.F. is the recipient of a Ruth L. Kirschstein National Research Service Award. The database and strain collection information is accessible at <http://yeastgfp.ucsf.edu>.

Authors' contributions Strain construction and analysis was performed by W.-K.H. and J.V.F., bioinformatics analysis and database management by L.C.G., with additional database management by A.S.C., and oligonucleotide primer design by R.W.H.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to E.K.O. (oshea@biochem.ucsf.edu).

Three-dimensional magnetic field topology in a region of solar coronal heating

S. K. Solanki¹, A. Lagg¹, J. Woch¹, N. Krupp¹ & M. Collados²

¹Max-Planck-Institut für Aeronomie, 37191 Katlenburg-Lindau, Germany

²Instituto de Astrofísica de Canarias, La Laguna, Tenerife, Spain

Flares and X-ray jets on the Sun arise in active regions where magnetic flux emerges from the solar interior and interacts with the ambient magnetic field^{1,2}. The interactions are believed to occur in electric current sheets separating regions of opposite magnetic polarity. The current sheets located in the corona or upper chromosphere have long been thought to act as an important source of coronal heating^{3–6}, requiring their location in the corona or upper chromosphere. The dynamics and energetics of these sheets are governed by a complex magnetic field structure that, until now, has been difficult to measure. Here we report the determination of the full magnetic vector in an interaction region near the base of the solar corona. The observations reveal two

magnetic features that characterize young active regions on the Sun: a set of rising magnetic loops and a tangential discontinuity of the magnetic field direction, the latter being the observational signature of an electric current sheet. This provides strong support for coronal heating models based on the dissipation of magnetic energy at current sheets.

The analysed observations were recorded on 13 May 2001 with the Tenerife Infrared Polarimeter (TIP)⁷ at the German Vacuum Tower Telescope (VTT) at the Spanish observatory of Izaña, Tenerife. We used observations in the He I 1083.0 nm multiplet^{8–11} to create maps of an emerging flux region in all four Stokes parameters *I*, *Q*, *U* and *V*, completely describing the polarized states of light. From the measured Stokes vector we derived the line-of-sight velocities, the strength of the magnetic field, its inclination to the solar surface normal and its direction in the horizontal plane (that is, the azimuth angle) in the photosphere and upper chromosphere. These parameters are plotted in Fig. 1. Within the scanned region the magnetic field changes its polarity from outward (away from the Sun, inclination angle close to 0°) in the East (left) to the opposite polarity in the West. On a scale larger than 10×10^6 m the shape and position of the boundary between the two polarities is similar in both layers. On smaller scales there are considerable deviations, with a strongly undulating boundary in the photosphere, including localized patches of opposite-polarity fields

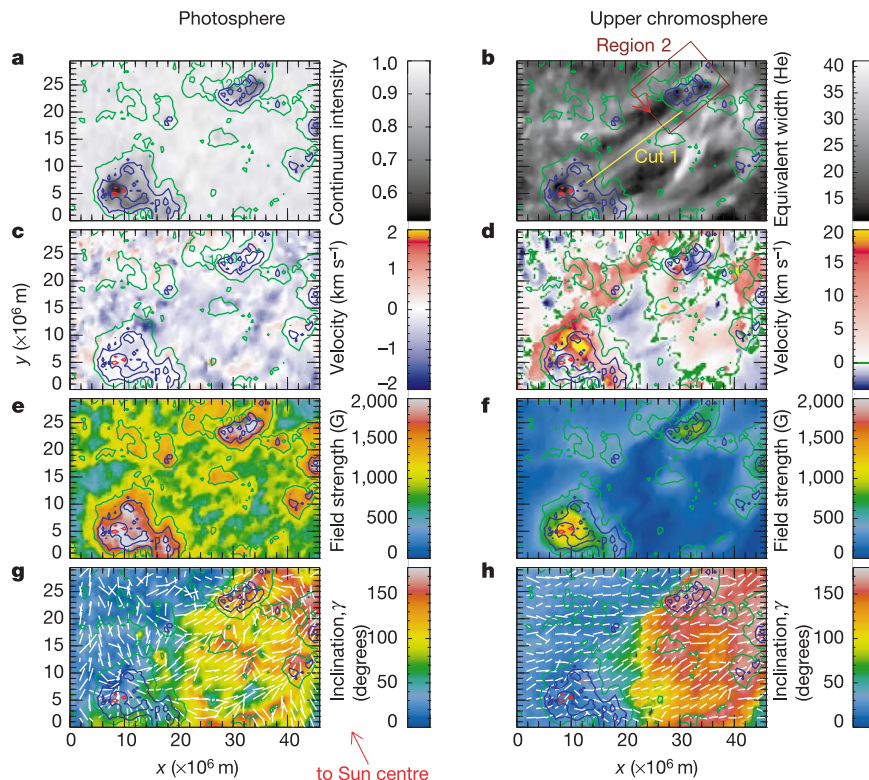


Figure 1 Atmospheric parameters of an emerging flux region (NOAA active region 9451, 33°W, 22°S). Displayed are from top to bottom continuum intensity (**a**) and the equivalent width of He I 1083.0 nm (**b**), the line-of-sight velocity, magnetic field strength and zenith angle (colour scale) with overplotted azimuthal direction (white lines). The photospheric maps (**a**, **c**, **e**, **g**) are derived from the measured Stokes profile of the Si I 1082.7 nm line (effective Landé factor $g_{\text{eff}} = 1.5$), the upper chromospheric maps (**b**, **d**, **f**, **h**) from the He I triplet at 1082.909 nm ($g_{\text{eff}} = 2.0$), 1083.025 nm ($g_{\text{eff}} = 1.75$) and 1083.034 nm ($g_{\text{eff}} = 0.875$). Positive values of the velocity denote downflows. The contour lines in each panel indicate photospheric magnetic field strength for reference (drawn at 1,200, 1,600, 2,000, 2,400 G). The pixel area is $0.39 \times 0.4''$ and the complete scan, recorded in 146 steps within 12 minutes, provides a field-of-view of $(46 \times 10^6) \times (30 \times 10^6)$ m²

on the Sun. Owing mainly to turbulence in the Earth's atmosphere the true spatial resolution is approximately $1.5''$. The Si I 1082.7 nm line, which is formed under local thermal equilibrium (LTE) conditions, was analysed by applying a response-function based inversion to the Stokes profiles^{17,18}. The model atmosphere consists of two components: a magnetic component for which temperature, the magnetic vector and line-of-sight velocity are deduced, and a field-free component corresponding to a quiet Sun model atmosphere¹⁹. The He I triplet, which has a complex non-LTE line formation²⁰ but is nearly optically thin²¹, was analysed by applying the Unno–Rachowski solution to describe the individual Zeeman components of each member of the triplet to the data²² combined with a simple implementation of the Hanle effect based on recent developments²³. Further details of the analysis are given elsewhere²⁴.

within the large-scale unipolar regions (visible in Fig. 1g). The magnetic structure is far more ordered in the upper chromosphere. This supports the view that whereas the Si I line basically samples a slab in the atmosphere (the photosphere) the He I triplet, being optically thin, gets contributions from all heights at which sufficient quantities of both neutral helium and ionizing radiation are present. Thus, at different locations within the emerging flux region the He I line is formed at different heights, following the corrugated coronal boundary.

The smooth change of the inclination angle from the upper right to the lower left (of Fig. 1h) deduced from the He triplet (that is, between the two newborn pores or continuum dark structures), suggests that the He I emission originates from loops connecting the two opposite photospheric polarities. This is supported by the fact that the field appears to be following the elongated filaments connecting the two pores seen in Fig. 1b. While a loop is still emerging it carries gas that has not yet been heated to coronal temperatures. Such gas is then visible in radiation absorbed by chromospheric spectral lines such as hydrogen H α and He I 1083.0 nm.

This interpretation is supported by Fig. 2, in which the components of the magnetic vector and the line-of-sight velocity obtained from He I along the yellow line marked 'Cut 1' in Fig. 1b are plotted. The remarkably smooth variation of the magnetic parameters along this cut is quite evident, with the strength of the

field decreasing steadily as it becomes increasingly horizontal. By contrast, the velocity increases from the middle towards both ends. This behaviour is very much like what one would expect when tracing the magnetic vector and the line-of-sight velocity along an emerging loop and projecting these quantities onto a plane. The cut shown in Fig. 2 suggests that the complete freshly emerged loops exhibit He I absorption (and not just their footpoints). These loops have not yet been in the solar atmosphere sufficiently long to let the relatively cool gas at chromospheric temperatures drain out, nor for it to have been heated to higher temperatures. Because the loops are still rising and draining (see Fig. 2c, showing a slow upflow in the centre and rapid, nearly sonic downflows near the edges), the structure is dynamic, so that He I absorption can occur at heights considerably larger than the scale height of chromospheric gas in a static situation.

Using simple criteria it is now possible to reconstruct the full three-dimensional structure of a magnetic loop in the solar atmosphere. The employed approach builds on the regular variation of the magnetic vector and the velocity field across the neutral line of the young active region (as exemplified by Fig. 2). The reconstruction starts by choosing a random pixel and identifying the two neighbouring pixels to which the magnetic azimuth points. If the field measured in these neighbouring pixels matches the field in the initial pixel in direction and strength, then they may belong to the same field line if the additional constraint is satisfied that the field

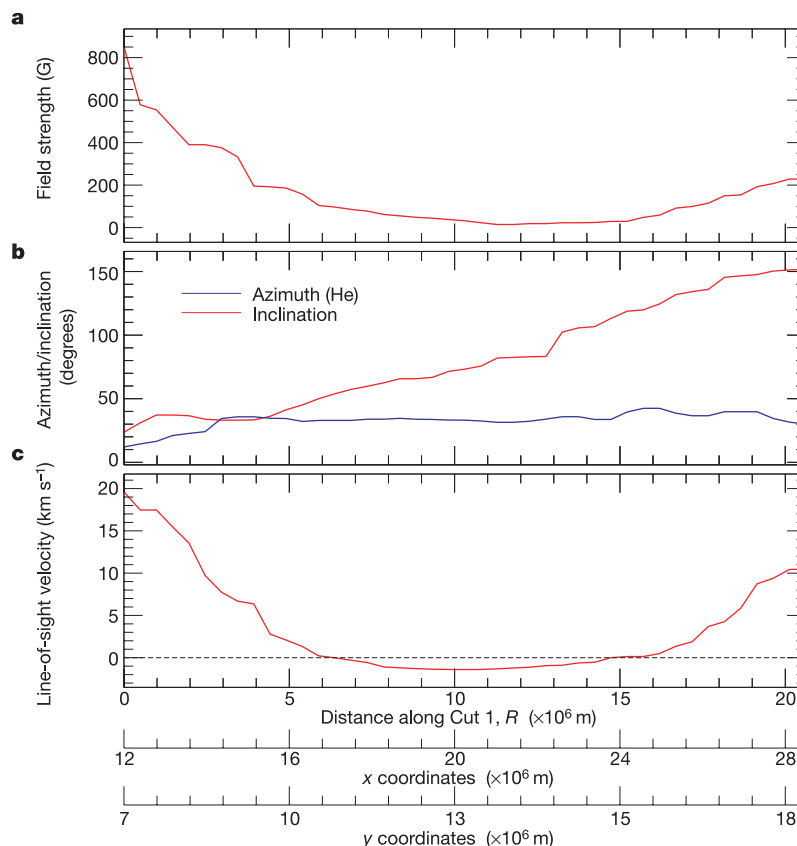


Figure 2 Magnetic and velocity field parameters obtained from the He triplet along 'Cut 1' in Fig. 1. For completeness, the horizontal axis lists the distance along the cut, R (top row of numbers), as well as the x and y coordinates taken from Fig. 1. The magnetic field strength decreases from a maximum value of 840 G at the edge to 20 G at the centre (a). The field is almost perpendicular to the solar surface near the edge of the cut (20° and 150° inclination) and is horizontal in the centre (inclination angle 90°, red curve in b). In this specific cut the azimuth stays almost constant between 30° and 40° within the

achieved accuracy over almost the whole length (b). This azimuth implies that the field follows the geometric angle of 'Cut 1'. The plotted magnetic profile is strongly suggestive of a loop. The velocity profile (c) shows the typical signature of freshly emerging magnetic flux—an upflow region in the centre (around the loop top) with velocities of 2 km s⁻¹ and a drainage of material with downflow velocities of up to 19.6 km s⁻¹ on the sides (loop footpoints). The horizontal dotted line represents gas at rest.

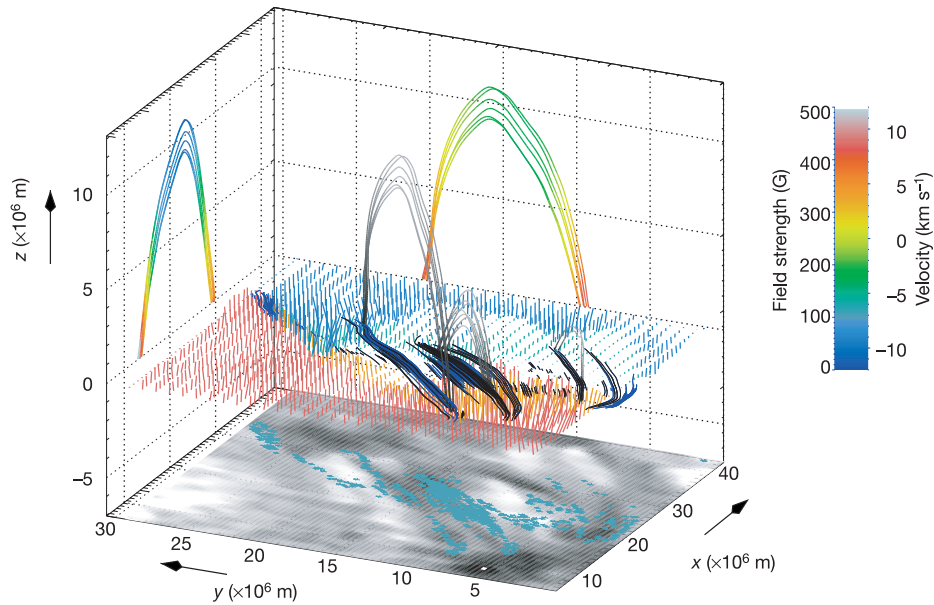


Figure 3 Reconstructed magnetic loops in the emerging flux region. The He I equivalent width map at the bottom is overlaid by a magnetic polarity and connectivity map over which examples of reconstructed loops are plotted. The derived azimuth and inclination angles of the magnetic field in each of the map's pixels are used to trace magnetic field lines. Heavy black lines are projections onto a fixed height of loops fulfilling all the criteria outlined in the text. Gratifyingly, these loops trace the He I equivalent width pattern (shown at the bottom of the plot). Some representative loop field lines are also plotted and for one of the plotted loops they are projected onto the x - z and y - z planes. The colour-coding of the loop projections represents the line-of-sight (basically vertical) velocity (y - z plane) and the magnetic field strength (x - z plane). The loops are rooted in areas with downflowing

material, while in the apex upflowing material is observed. This observation, combined with the frozen-in horizontal magnetic field, implies that the whole loop is rising. The magnetic field strength decreases with height in both legs from about 390 and 500 G at the two chromospheric footpoints to below 50 G at the apex. Magnetic field vectors at the footpoints of the field lines are shown to visualize the overall magnetic field geometry and strength. Field lines out of (or into) the plane parallel to the solar surface are coded in red and orange (or blue and green), with orange and green marking field lines that close within the tracing box, that is, for which both footpoints lie within the observed area and He I emission is seen from the whole length of the loop.

strength decreases with height. The height sampled in the neighbouring pixel relative to that in the original pixel is determined by the magnetic inclination and the size of the pixel. If the field is stronger or points in a distinctly different direction then we expect that the radiation in the two pixels is coming from different loops. Thus, starting from a given pixel, we follow a field line in both directions until one of the above criteria is no longer satisfied. If the end-points of such a field line lie at roughly the same height, have opposite polarities and if neighbouring pixels of one end-point map to neighbouring pixels of the other end-point, then we identify the traced bundle of field lines with a loop. The loops recovered in this manner are three-dimensional structures, because we have information on the full magnetic vector. Examples of reconstructed loops are plotted in Fig. 3. The loops show an asymmetry, such as in the field strength between the two footpoints. The loops fan out as the field strength drops with height (nearly exponentially, with a scale height for the largest loops of $4\text{--}5 \times 10^6$ m). This is in contrast to the very thin, fibril-like loops observed by the Transition Region and Coronal Explorer (TRACE)¹², suggesting that the emission recorded by TRACE selectively highlights a few field lines. Using the deduced loop geometry we determine that the downflows along the loop legs correspond to freely falling material¹³.

The maps made in the He I lines reveal locations of strong gradients in the magnetic vector. The strongest gradient lies in the box marked 'Region 2' in Fig. 1. A blow-up of the field strength in this box is displayed in Fig. 4. Clearly, the field strength drops abruptly along the neutral line, forming a deep valley whose width corresponds closely to the estimated spatial resolution of $1.5''$. Obviously, this feature has not been fully resolved. The strength of the computed electric current is represented by the colours. Figure 4 strongly suggests that we are dealing here with a current

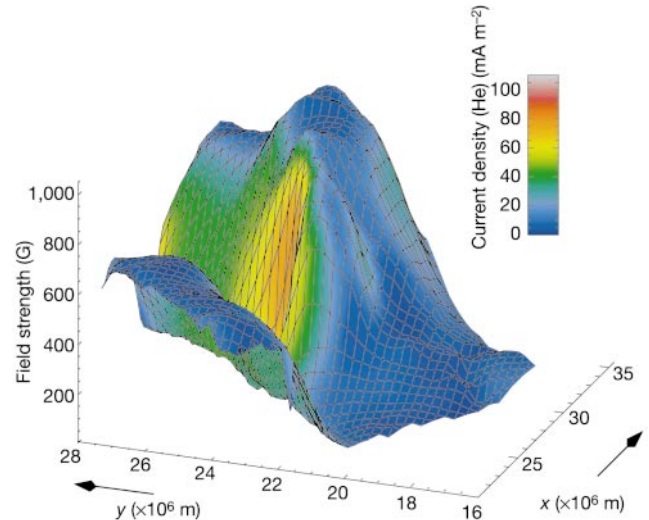


Figure 4 Representation of the electric current sheet (See 'Region 2' in Fig. 1 for the location within the full scanned region). The elevation of the meshed surface is proportional to the magnetic field strength in the upper chromosphere. A narrow valley of low magnetic field values separates two areas of opposite magnetic polarity, so that an electric current flows along this boundary parallel to the solar surface. The colour-coding in the figure indicates the current density of this current sheet calculated using Ampère's law. The maximum value of approximately 90 mA m^{-2} in the region of the largest magnetic field gradient perpendicular to the current sheet represents only a lower limit of the actual current flowing in this chromospheric current sheet. The width of the valley in the magnetic field strength corresponds to the spatial resolution of 1×10^6 m, limited by turbulence in the Earth's atmosphere, so that the horizontal gradient of the magnetic field is underestimated.

sheet. The maximum value of 90 mA m^{-2} is a lower limit, because we do not resolve the current sheet.

This is, to our knowledge, the first direct evidence of an electric current sheet in the upper solar atmosphere, whose presence has long been predicted. Currents are usually deduced from photospheric vector magnetograms¹⁴. Outside sunspots the interpretation of the photospheric currents is not straightforward, because there the field is highly filamented and unresolved. In the upper chromosphere it has expanded to fill all of space, so that an interpretation in terms of a current sheet is reasonable. Current sheets are thought to be formed spontaneously when the footprints of the Sun's magnetic field are shuffled around^{3,4,15} or when new magnetic flux emerges into the solar atmosphere. Our observations demonstrate that vector spectropolarimetry of the He I 1083.0 nm line allows such current sheets to be detected, thus providing the possibility of observationally testing their importance for the triggering of flares or for coronal heating¹⁶. The presence of such a clear signature in the very first magnetic vector map made with the He I line provides strong support for coronal heating theories based on magnetic energy dissipation at current sheets. □

Received 21 March; accepted 2 September 2003; doi:10.1038/nature02035.

1. Yokoyama, T. & Shibata, K. Magnetic reconnection as the origin of X-ray jets and H-alpha surges in the Sun. *Nature* **375**, 42–44 (1995).
2. Lites, B. W., Skumaniah, A. & Martinez Pillet, V. Vector magnetic fields of emerging solar flux. I. Properties at the site of emergence. *Astron. Astrophys.* **333**, 1053–1068 (1998).
3. Parker, E. N. Magnetic neutral sheets in evolving fields—Part One—General theory. *Astrophys. J.* **264**, 635–641 (1983).
4. Parker, E. N. Magnetic neutral sheets in evolving fields—Part Two—Formation of the solar corona. *Astrophys. J.* **264**, 642–647 (1983).
5. Parker, E. N. Nanoflares and the solar X-ray corona. *Astrophys. J.* **330**, 474–479 (1988).
6. Ulmschneider, P., Priest, E. R. & Rosner, R. (eds) *Mechanism of Chromospheric and Coronal Heating* Ch. 4 (Springer, Berlin/Heidelberg/New York, 1991).
7. Martínez Pillet, V. et al. in *ASP Conf. Ser. 183. High Resolution Solar Physics: Theory, Observation, and Techniques* 264 (Astronomical Society of the Pacific, San Francisco, California, 1999).
8. Harvey, J. & Hall, D. in *IAU Symp. 43. Solar Magnetic Fields* (ed. Howard, R.) 279–288 (Dordrecht, Reidel, 1971).
9. Penn, M. J. & Kuhn, J. R. Imaging spectropolarimetry of the He I 1083 nanometer line in a flaring solar active region. *Astrophys. J.* **441**, L51–L54 (1995).
10. Rüedi, L., Solanki, S. K. & Livingston, W. C. Infrared lines as probes of solar magnetic features. X. He I 10830 Å as a diagnostic of chromospheric magnetic fields. *Astron. Astrophys.* **293**, 252–262 (1995).
11. Rüedi, L., Keller, C. U. & Solanki, S. K. Measurements of the full Stokes vector of He I 10830 Å. *Sol. Phys.* **164**, 265–275 (1996).
12. Handy, B. N. et al. The transition region and coronal explorer. *Sol. Phys.* **187**, 229–260 (1999).
13. Schmidt, W., Muglach, K. & Knölker, M. Free-fall downflow observed in He I 1083.0 nm and Hβ. *Astrophys. J.* **544**, 567–571 (2000).
14. Zirin, H. & Wang, H. Narrow lanes of transverse magnetic field in sunspots. *Nature* **363**, 426–428 (1993).
15. Galsgaard, K. & Longbottom, A. W. Formation of solar prominences by flux convergence. *Astrophys. J.* **510**, 444–459 (1999).
16. Gudiksen, B. V. & Nordlund, Å. Bulk heating and slender magnetic loops in the solar corona. *Astrophys. J.* **572**, L113–L116 (2002).
17. Ruiz Cobo, B. & del Toro Iniesta, J. C. Inversion of Stokes profiles. *Astrophys. J.* **398**, 375–385 (1992).
18. Frutiger, C., Solanki, S. K., Fligge, M. & Bruls, J. H. M. J. Properties of the solar granulation obtained from the inversion of low spatial resolution spectra. *Astron. Astrophys.* **358**, 1109–1121 (2000).
19. Kurucz, R. L. 'Finding' the 'Missing' solar ultraviolet opacity. *Rev. Mexicana Astron. Astrofis.* **23**, 181–186 (1992).
20. Avrett, E. H., Fontenla, J. M. & Loeser, R. in *IAU Symp. 154. Infrared Solar Physics* (ed. Rabin, D. M.) 35–47 (Kluwer Academic, Dordrecht, 1994).
21. Giovanelli, R. G. & Hall, D. The helium 10830 Å line in the undisturbed chromosphere. *Sol. Phys.* **52**, 211–228 (1977).
22. Rachkowsky, D. N. The reduction for anomalous dispersion in the theory of the absorption line formation in a magnetic field [in Russian]. *Izv. Krym. Astrofiz. Obs.* **37**, 56–61 (1967).
23. Trujillo Bueno, J., Landi Degl'Innocenti, E., Collados, M., Merenda, L. & Manso Sainz, R. Selective absorption processes as the origin of puzzling spectral line polarization from the Sun. *Nature* **415**, 403–406 (2002).
24. Lagg, A., Woch, J., Krupp, N. & Solanki, S. K. Retrieval of the full magnetic vector with the He I multiplet at 1083 nm. *Astron. Astrophys.* (submitted).

Acknowledgements M.C. acknowledges support from the Spanish Ministerio de Ciencia y Tecnología. We thank the Kiepenheuer Institut in Freiburg, Germany, for observing support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.K.S. (solanki@limpi.mpg.de).

The speed of information in a 'fast-light' optical medium

Michael D. Stenner¹, Daniel J. Gauthier¹ & Mark A. Neifeld²

¹Duke University, Department of Physics, and The Fitzpatrick Center for Photonics and Communication Systems, Durham, North Carolina 27708, USA

²Department of Electrical and Computer Engineering, The Optical Sciences Center, University of Arizona, Tucson, Arizona 85721, USA

One consequence of the special theory of relativity is that no signal can cause an effect outside the source light cone, the space-time surface on which light rays emanate from the source¹. Violation of this principle of relativistic causality leads to paradoxes, such as that of an effect preceding its cause². Recent experiments on optical pulse propagation in so-called 'fast-light' media—which are characterized by a wave group velocity v_g exceeding the vacuum speed of light c or taking on negative values³—have led to renewed debate about the definition of the information velocity v_i . One view is that $v_i = v_g$ (ref. 4), which would violate causality, while another is that $v_i = c$ in all situations⁵, which would preserve causality. Here we find that the time to detect information propagating through a fast-light medium is slightly longer than the time required to detect the same information travelling through a vacuum, even though v_g in the medium vastly exceeds c . Our observations are therefore consistent with relativistic causality and help to resolve the controversies surrounding superluminal pulse propagation.

The speed of a light pulse travelling through an optical material is not precisely defined, because any pulse comprises a collection of elementary sinusoidal waveforms, each with a distinct frequency ω . Each constituent sinusoid travels at a well-defined velocity known as the phase velocity $v_p = c/n(\omega)$, where $n(\omega)$ is the refractive index of the optical material. Approximate theories of optical pulse propagation predict that the peak travels at the group velocity $v_g = c/(n + \omega dn/d\omega|_{\omega=\omega_0}) = c/n_g$, where n_g is the group index and ω_0 is the central frequency of the wavepacket⁶.

We refer to the quantity $dn/d\omega$ as the dispersion of an optical material. For typical optical materials, there exist narrow spectral regions where $n(\omega)$ is a decreasing function of frequency (that is, $dn/d\omega < 0$), resulting in a condition known as anomalous dispersion⁷. When ω_0 is within such a region, n_g can be less than one and can even become negative when the anomalous dispersion is large. This results in 'fast light', for which it is possible that the peak of a light pulse may exit the optical material before it passes through the entrance face⁸. The amount of fast-light pulse advancement is largest when v_g is negative and near zero (n_g large and negative).

The possibility of superluminal group velocities ($v_g > c$ or $v_g < 0$) was such a concern to researchers around 1910 that several conference sessions were devoted to the topic⁹. Based on these discussions, Sommerfeld demonstrated theoretically that the velocity of the front of a square-shaped pulse propagating through any medium is identically equal to c and hence relativistic causality is preserved¹⁰. In a follow-up study, Brillouin suggested that the group velocity is not physically meaningful when the dispersion is anomalous because the pulse becomes severely distorted¹¹. More recent research investigating the propagation of smooth-shaped pulses has shown that this conclusion is not justified, leading to renewed controversy^{8,12–20}.

Another outcome of the discussions in the early 1900s, as recounted in the preface and first chapter of the book by Brillouin⁹, was a reformulation of the fundamental postulate of the special theory of relativity. This reformulation states that, rather than limiting the speed of an 'object', it is the information velocity v_i that is limited by c . Unfortunately, there is no agreed-upon definition of the information velocity².

In our experiment, we use a fast-light medium that exploits the spectral region of anomalous dispersion between two closely spaced amplifying resonances^{15,20} realized by creating large atomic coherence²¹ in a laser-driven potassium vapour²², as shown in Fig. 1a. We obtain larger pulse advancement for a smooth gaussian-shaped pulse, as shown in Fig. 1b, in comparison to the experiment of ref. 15, by increasing the gain and hence the size of the anomalous dispersion. The larger advancement relative to the pulse width obtained in our experiment makes it easier to distinguish the different velocities describing pulse propagation. From this data, we infer that $n_g = -19.6 \pm 0.8$, indicating that we are operating in the highly superluminal regime.

Measuring v_i requires an understanding of the fundamental mechanism for information encoding and detection. Garrison *et al.*² propose that new information is encoded on an optical pulse by creating a point that is non-analytic (for example, a discontinuity in the pulse amplitude or its derivatives) and that this point always travels at c regardless of the value of the other velocities associated with the pulse^{2,5}. Essentially, they have generalized Sommerfeld's concept of the front velocity to a non-analytic point of the pulse amplitude, where the front of a square-

shaped pulse is an example of a point of non-analyticity. These workers^{2,5} suggest that the point of non-analyticity is the only part of the pulse representing new information because measurements of the early part of the pulse cannot be used to predict anything about the part of the pulse arriving after the point of non-analyticity, and hence v_i equals the speed of a point of non-analyticity. For counter-views, see refs 2, 4, 23 and 24. We note that some aspects of their proposal have been verified using electronic circuits where no propagating waves are involved, and hence only issues of causality, but not relativistic causality, can be tested^{25–27}.

To enhance our ability to estimate the location of this non-analytic point in the presence of noise, we use two optical pulses that are initially identically gaussian-shaped, which allows us smoothly to turn on the pulse amplitude to a level above the noise floor of our detection electronics and to monitor the fast-light pulse advancement. Near the peak of the gaussian function and at the same moment for both symbols, we switch the amplitude of the gaussian function to a high (1) or low (0) value for the remainder of the pulse. The moment when a decision is made to switch between the symbols corresponds to the point of non-analyticity. Note that this transition is smoothed out by the finite response time of the optical switch.

The location of the point of non-analyticity is determined by detecting the arrival of new information using a receiver that can distinguish between symbols to a desired level of certainty, characterized by the bit error rate (BER). Before the arrival of the point of non-analyticity at the detector, we expect no detected information, corresponding to a BER of 1/2. Once the point of non-analyticity propagates past the detector, the received information will grow smoothly from zero and the BER drops. A symbol is considered to be detected when the BER falls below some threshold. Hence, the detection time of information is later than the time when information is first available at the detector, even for pulses propagating through vacuum. This detection latency Δt depends on the charac-

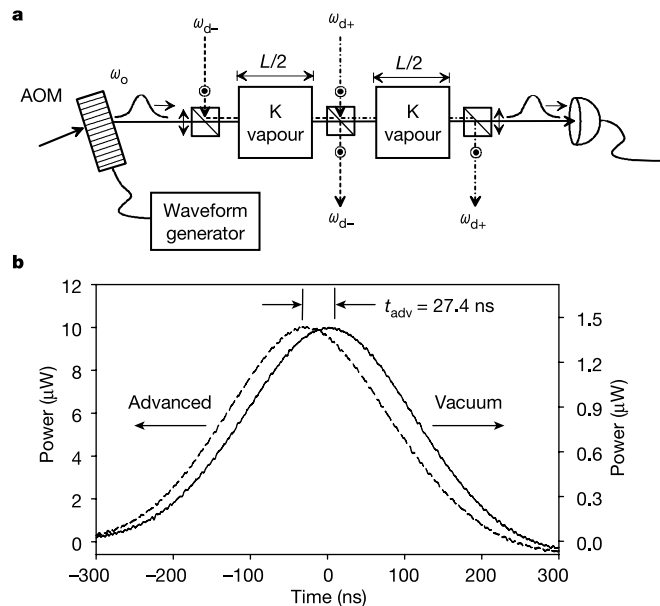


Figure 1 Fast-light pulse propagation. **a**, Experimental set-up. The potassium vapours are contained in two uncoated Pyrex cells of length $L/2 = 20 \text{ cm}$ (to suppress unwanted parametric instabilities²²) and heated to obtain an atomic number density of $4.5 \times 10^{11} \text{ atoms cm}^{-3}$. Linearly polarized coherence-preparation laser beams (frequencies ω_{d-} and ω_{d+}) are combined with the linear and orthogonally polarized pulses using polarizing beam splitters. The pulses are detected by a photoreceiver with a 25 kHz–125 MHz bandwidth. The coherence preparation beams are adjusted with ω_{d-} set at 1.36 GHz to the high-frequency side of the centre of the potassium $4S_{1/2} \leftrightarrow 4P_{1/2}$ transition and $\omega_{d+} - \omega_{d-} = 23 \text{ MHz}$, chosen to optimize the pulse advancement using procedures similar to those discussed in refs 15, 18 and 22. The pulses are generated by passing a continuous-wave laser beam through an acousto-optic modulator (AOM) driven by a computer-controlled arbitrary waveform generator. The time origin has been set arbitrarily to coincide with the peak of this pulse. **b**, The solid line shows the temporal evolution of a 263.4-ns-long (full-width at half-maximum) pulse propagating through the cells when the lasers are tuned far from the atomic resonance and hence the vapour-cell portion of the path is equivalent to vacuum. The dashed line shows the observed fast-light pulse advancement for a smooth pulse shape when the coherence-preparation laser is tuned near the atomic resonance and ω_o is set between the gain resonances. The peak of the pulse is advanced by $t_{\text{adv}} = 27.4 \text{ ns} \pm 1.1 \text{ ns}$, corresponding to a relative pulse advancement of 10.4%. Using $t_{\text{adv}} = L/c - L/v_g$ with $L/c = 1.3 \text{ ns}$, we find $v_g/c = -0.051 \pm 0.002$. Careful inspection of the fast-light pulse reveals that it has been compressed by 1.9%, which is due primarily to the frequency dependence of the gain^{18,19}.

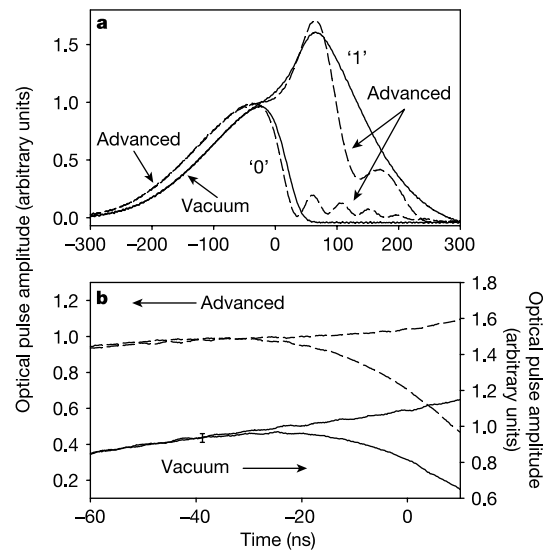


Figure 2 Transmitting information-encoded optical pulses through a fast-light medium. **a**, Transmitting '0' and '1' through the fast-light medium (dashed line) and vacuum (solid line). Each symbol is transmitted separately through the medium and vacuum, where each curve is an average of 50 pulses. **b**, High-resolution plot of part **a** with an offset for clarity. The amplitude of the advanced and vacuum pulses have been scaled so that their heights would be the same if a gaussian pulse propagated through the system, as in Fig. 1b. The error bar indicates the typical standard deviation of the pulse amplitudes. From **a**, it is seen that the fast-light medium advances the early part of the pulses during the smooth turn on, identically to that observed for the full gaussian-shaped pulses shown in Fig. 1b. Most important is the observation that both symbols are the same for early times so that it is not possible to distinguish between them. Hence, no information can yet be conveyed to a receiving party at the end of the communication channel.

teristics of the medium through which the pulses propagate, the shape of the symbols, the detection algorithm, noise²⁴ in the detection process (including quantum noise^{28,29}), and the BER threshold. It increases as the signal-to-noise ratio decreases because it takes longer for the receiver to achieve the same BER. Achieving the limit $\Delta t \rightarrow 0$ requires the use of optimal symbol shapes and detection algorithms, and infinite energy in the optical pulse so that the signal-to-noise ratio of the detected waveform approaches infinity. Although it is possible to estimate Δt for a specific experimental apparatus, it cannot be measured directly because it requires measuring the point of non-analyticity. A crucial aspect of our experiment is to make Δt as small as possible and to make it as similar as possible for both vacuum and advanced pulses.

Figure 2 shows the propagation of both symbols through the fast-light medium and vacuum. From a simple visual inspection of the data, we see that the time where it is possible to first distinguish between the two symbols for the advanced pulses is slightly later than the time where it is possible to first distinguish between the same two symbols propagating through vacuum. In addition, it is seen that the manner in which the average symbol waveforms separate for the vacuum and advanced case are only slightly different, so that the detection latency times (denoted by Δt_{vac} and Δt_{adv} respectively) should be similar.

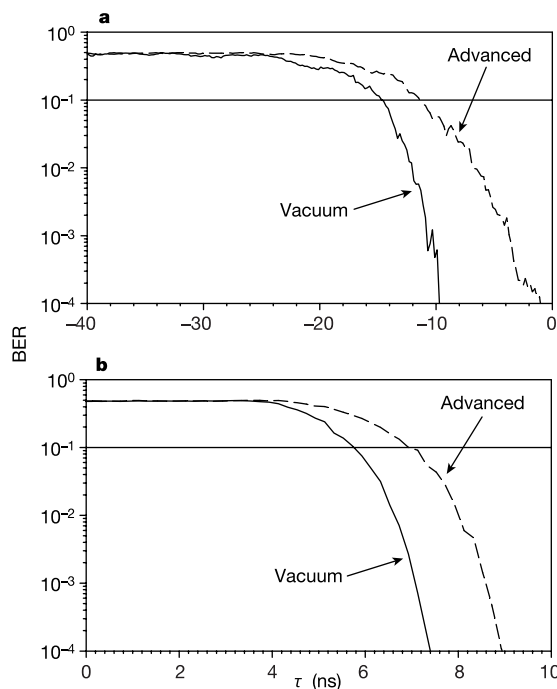


Figure 3 Detecting the arrival of new information. Shown is the BER as a function of the upper limit of the integration time τ for the vacuum (solid line) and advanced (dashed line) pulses. The horizontal line indicates the detection threshold. **a**, Experimental observations. The time origin has been selected arbitrarily. **b**, Theoretical predictions based on Maxwell's equations. The time origin corresponds to the moment when the point of non-analyticity first arrives at the detector. The BER is determined using the following matched-filter procedure. For each symbol (0 and 1), there are 50 pulse waveforms. Reference waveforms, denoted by $R_0(t)$ and $R_1(t)$, are generated by averaging 49 pulse waveforms and the integral $D(\tau) = I_0(\tau) - I_1(\tau)$ is determined for each pulse, where the waveform being detected is not included in the respective reference waveform. Here: $I_j(\tau) = \int_{t_s}^{t_s+\tau} x(t)R_j(t)dt/\alpha_j(\tau_\alpha)N_j(\tau)$ (for $j = 0, 1$), where $x(t)$ is an individual pulse waveform, t_s is an integration start time chosen arbitrarily at a time in the early beginning of the pulse, $N_j(\tau) = \int_{t_s}^{t_s+\tau} R_j^2(t)dt$, $\alpha_j(\tau_\alpha) = \int_{t_s}^{t_s+\tau_\alpha} x^2(t)dt/N_j(\tau_\alpha)$, and τ_α is a normalization integration time chosen arbitrarily on the leading edge of the gaussian pulse before the point where the symbols separate. For each symbol type, the probability density of D is fitted to a gaussian distribution normalized to an area of 1/2 and the overlapping area of these two gaussian distributions is the BER.

To quantify our results, we determine the BER for the vacuum (Fig. 3a, solid line) and advanced (dashed line) pulse pairs using an integrate-and-dump matched filter technique. The BER is high for final observation times in the range between -40 and -25 ns, during which the pulse amplitudes are large (see Fig. 2a). Hence, even though the signal-to-noise ratio for a single pulse is high at these times, the pulses are not yet distinguishable and no information is detected. Placing the detection threshold at $\text{BER} = 0.1$, chosen to keep Δt_{vac} and Δt_{adv} small, we determine the detection time for vacuum (advanced) pulse pairs T_{vac} (T_{adv}) and the difference in detection times $T_i = T_{\text{adv}} - T_{\text{vac}}$. The time difference is approximately constant for BER values around 0.1; its average value in the range of BERs between 0.08 and 0.2 is equal to 3.2 ± 1.5 ns. Our observations demonstrate that the information detection time for pulses propagating through the fast-light medium is longer than the detection time for the same information propagating through vacuum, even though the group velocity is in the highly superluminal regime for the fast-light medium.

From our direct measurement of T_i , we do not know whether an observed difference between the detection times is due to changes in the detection latencies or differences in the information velocities for vacuum ($v_{i,\text{vac}}$) and the fast-light medium ($v_{i,\text{adv}}$). The relation among these quantities is given by:

$$T_i = (L/v_{i,\text{adv}} - L/v_{i,\text{vac}}) + (\Delta t_{\text{adv}} - \Delta t_{\text{vac}}) \quad (1)$$

To gain some insight about the importance of detection latency, we analyse a mathematical model, based on Maxwell's equations, that describes approximately the generation, propagation, and detection of our symbols. Consistent with previous research^{2,5}, this model predicts that $v_{i,\text{adv}} = v_{i,\text{vac}} = c$, and hence T_i is completely determined by $(\Delta t_{\text{adv}} - \Delta t_{\text{vac}})$. Using the same matched-filtering approach, we determine the predicted BER as shown in Fig. 3b. We see that information is detected later for the advanced pulses than for the vacuum pulses, qualitatively similar to the experimental observations. We find that $T_i = 1.5 \pm 0.5$ ns, where the error only accounts for statistical uncertainty in the BER determination. The fact that $T_i \neq 0$ demonstrates that subtle changes in the shape of the symbols after information has been encoded give rise to substantial changes in the detection latency. The predicted value of T_i might be smaller than the observed time owing to our assumption that the fast-light medium does not change the noise properties of the optical pulses²⁸.

Using the model prediction for $(\Delta t_{\text{adv}} - \Delta t_{\text{vac}})$ in equation (1) and taking $v_{i,\text{vac}} = c$, we find that $v_{i,\text{adv}} = 0.4(+0.7 - 0.2)c$. Thus, our observations are consistent with the special theory of relativity even for a medium where v_g is highly superluminal, demonstrating that the peak of the advanced pulse at the exit face of the medium (see Fig. 1b) is not causally connected to the peak at the entrance face². Because our analysis makes no assumptions about the sources of noise in the encoding, transmission and decoding process, our general experimental approach and conclusions should hold even in the limit where quantum fluctuations are dominant^{3,28,29}. □

Received 11 July; accepted 20 August 2003; doi:10.1038/nature02016.

1. Jackson, J. D. *Classical Electrodynamics* 3rd edn, Sec. 11.3.C (John Wiley & Sons, New York, 1999).
2. Garrison, J. C., Mitchell, M. W., Chiao, R. Y. & Bolda, E. L. Superluminal signals: causal loop paradoxes revisited. *Phys. Lett. A* **245**, 19–25 (1998).
3. Boyd, R. W. & Gauthier, D. J. in *Progress in Optics* (ed. Wolf, E.) Vol. 43, 497–530, Ch. 6 (Elsevier, Amsterdam, 2002).
4. Nimtz, G. & Gaebel, A. Basics of superluminal signals. *Ann. Phys.* **11**, 163–171 (2002).
5. Chiao, R. Y. & Steinberg, A. M. in *Progress in Optics* (ed. Wolf, E.) Vol. 37, 345–405, Ch. 6 (Elsevier Science, Amsterdam, 1997).
6. Agrawal, G. P. *Nonlinear Fiber Optics* 3rd edn, Ch. 1 (Academic, San Diego, 2001).
7. Born, M. & Wolf, E. *Principles of Optics* 7th edn, Sec. 2.3.4 (Cambridge Univ. Press, Cambridge, 1999).
8. Garrett, C. G. B. & McCumber, D. E. Propagation of a Gaussian light pulse through an anomalous dispersion medium. *Phys. Rev. A* **1**, 305–313 (1970).
9. Brillouin, L. *Wave Propagation and Group Velocity* Ch. I (Academic, New York, 1960).
10. Sommerfeld, A. Über die fortpflanzung des liches in dispergierenden medien. *Ann. Phys.* **44**, 177–202 (1914); English translation available in Brillouin, L. *Wave Propagation and Group Velocity* Ch. II (Academic, New York, 1960).
11. Brillouin, L. Über die fortpflanzung des liches in dispergierenden medien. *Ann. Phys.* **44**, 203–240

- (1914); English translation available in Brillouin, L. *Wave Propagation and Group Velocity* Ch. III (Academic, New York, 1960).
12. Chu, C. & Wong, S. Linear pulse propagation in an absorbing medium. *Phys. Rev. Lett.* **48**, 738–741 (1982).
 13. Ségard, B. & Macke, B. Observation of negative velocity pulse propagation. *Phys. Lett.* **109**, 213–216 (1985).
 14. Peatross, J., Glasgow, S. A. & Ware, M. Average energy flow of optical pulses in dispersive media. *Phys. Rev. Lett.* **84**, 2370–2373 (2000).
 15. Wang, L. J., Kuzmich, A. & Dogariu, A. Gain-assisted superluminal light propagation. *Nature* **406**, 277–279 (2000).
 16. Akulshin, A. M., Cimmino, A. & Opat, G. I. Negative group velocity of a light pulse in cesium vapor. *Quantum Electron.* **32**, 567–569 (2002).
 17. Bigelow, M. S., Lepeshkin, N. N. & Boyd, R. W. Superluminal and slow light propagation in a room-temperature solid. *Science* **301**, 200–202 (2003).
 18. Cao, H., Dogariu, A. & Wang, L. J. Negative group delay and pulse compression in superluminal pulse propagation. *IEEE J. Sel. Top. Quantum Electron.* **9**, 52–58 (2003).
 19. Macke, B. & Ségard, B. Propagation of light-pulses at a negative group velocity. *Eur. Phys. J. D* **23**, 125–141 (2003).
 20. Steinberg, A. M. & Chiao, R. Y. Dispersionless, highly superluminal propagation in a medium with a gain doublet. *Phys. Rev. A* **49**, 2071–2075 (1994).
 21. Harris, S. E. Electromagnetically induced transparency. *Phys. Today* **50**, 36–42 (1997).
 22. Stenner, M. D. & Gauthier, D. J. Pump-beam-instability limits to Raman-gain-doublet “fast-light” pulse propagation. *Phys. Rev. A* **67**, 063801 (2003).
 23. Diener, G. Superluminal group velocities and information transfer. *Phys. Lett. A* **223**, 327–331 (1996).
 24. Wynne, K. Causality and the nature of information. *Opt. Commun.* **209**, 85–100 (2002).
 25. Mitchell, M. W. & Chiao, R. Y. Negative group delay and “fronts” in a causal system: An experiment with very low frequency bandpass filters. *Phys. Lett. A* **230**, 133–138 (1999).
 26. Nakanishi, T., Sugiyama, K. & Kitano, K. Demonstration of negative group delays in a simple electronic circuit. *Am. J. Phys.* **70**, 1117–1121 (2002).
 27. Solli, D., Chiao, R. Y. & Hickmann, J. M. Superluminal effects and negative group delays in electronics, and their applications. *Phys. Rev. E* **66**, 056601 (2002).
 28. Kuzmich, A., Dogariu, A., Wang, L. J., Milonni, P. W. & Chiao, R. Y. Signal velocity, causality, and quantum noise in superluminal light pulse propagation. *Phys. Rev. Lett.* **86**, 3925–3929 (2001).
 29. Caves, C. M. & Drummond, P. Quantum limits on bosonic communication rates. *Rev. Mod. Phys.* **66**, 481–537 (1994).

Acknowledgements M.D.S. and D.J.G. acknowledge discussions with M. Gehm and J. Thomas, the loan of an argon-ion pump laser from J. Thomas, and the financial support of the US National Science Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.G. (gauthier@phy.duke.edu).

Single-electron transistor of a single organic molecule with access to several redox states

Sergey Kubatkin¹, Andrey Danilov¹, Mattias Hjort², Jérôme Cornil^{2,3}, Jean-Luc Brédas^{2,3*}, Nicolai Stühr-Hansen⁴, Per Hedegård⁴ & Thomas Bjørnholm⁴

¹Department of Microtechnology and Nanoscience/MC2, University of Technology, Chalmers, Göteborg, Sweden

²Department of Chemistry, The University of Arizona, Tucson, Arizona 85721-0041, USA

³Laboratory for Chemistry of Novel Materials, Center for Research in Molecular Electronics and Photonics, University of Mons-Hainaut, Place du Parc 20, B-7000 Mons, Belgium

⁴Nano-Science Center (Department of Chemistry and Niels Bohr Institute), University of Copenhagen, Universitetsparken 5, DK-2100, Copenhagen, Denmark

* Present address: School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-040, USA

A combination of classical Coulomb charging, electronic level spacings, spin, and vibrational modes determines the single-electron transfer reactions through nanoscale systems connected to external electrodes by tunnelling barriers¹. Coulomb charging effects have been shown to dominate such transport in semi-

conductor quantum dots², metallic³ and semiconducting⁴ nanoparticles, carbon nanotubes^{5,6}, and single molecules^{7–9}. Recently, transport has been shown to be also influenced by spin—through the Kondo effect—for both nanotubes¹⁰ and single molecules^{8,9}, as well as by vibrational fine structure^{7,11}. Here we describe a single-electron transistor where the electronic levels of a single π -conjugated molecule in several distinct charged states control the transport properties. The molecular electronic levels extracted from the single-electron-transistor measurements are strongly perturbed compared to those of the molecule in solution, leading to a very significant reduction of the gap between the highest occupied molecular orbital and the lowest unoccupied molecular orbital. We suggest, and verify by simple model calculations, that this surprising effect could be caused by image charges generated in the source and drain electrodes resulting in a strong localization of the charges on the molecule.

We measured electrical transport at 4 K through a single *p*-phenylenevinylene oligomer, which has five benzene rings connected through four double bonds (OPV5, Fig. 1a). OPV5 was placed by chemical vapour deposition in a gap about 2 nm wide separating the source and drain electrodes of a single-electron transistor (SET) device (Fig. 1). The synthesis of OPV5 ((*E,E*)-1,4-bis{4-[(*E*)-4-(*tert*-butylthio)styryl]}benzene) was done by a Horner–Wadsworth–Emmons condensation¹² of tetraethyl 1,4-xylylenediphosphonate with *trans*-4-(*tert*-butylthio)styrylbenzaldehyde. The terminal thiols were protected with a *tert*-butyl group, which prevents chemical binding of the sulphur to the gold electrode and leads to a weak van der Waals contact between the molecule and the source and drain electrodes.

A planar gate electrode made of aluminium metal covered with aluminium oxide (~5 nm thick) was prepared on a chip of oxidized silicon. A shadow mask used to deposit the gold lead electrodes was defined on top of the gate by standard electron-beam lithography. The chip was then introduced into a vacuum chamber immersed in liquid helium. All subsequent procedures reported here were performed during a single vacuum cycle. First, two gold electrodes were deposited through a shadow mask by condensing gold vapour on the substrate held at 4.2 K. By using an oblique evaporation angle (Fig. 1b) together with *in situ* conductance measurements, we were able to fine-tune the tunnelling gap between the gold electrodes to a few nanometres¹³. As shown in Fig. 1b, for an inclined beam, a constriction is formed by two mask edges. If the tilt angle is high, there is no overlap between source and drain shadows. Reducing the tilt angle decreases the source–drain gap. We changed the tilt stepwise, narrowing the gap between the leads by 5 nm at each step. At each step, a 2-nm-thick film of gold was deposited through the mask and the sample was checked for non-zero tunnelling conductance (this procedure of nanometre-scale gap fabrication has been successfully tested for other metals¹³). Eventually, we fabricated two self-aligned and self-sharpened gold electrodes with a tunnelling gap of a few M Ω between them. By annealing the sample up to 100 K, we increased the gap resistance to a few G Ω , which corresponds to a tunnelling gap width of roughly 2 nm (ref. 14). Annealed samples did not show any gate dependence of the tunnelling conductance nor any peculiarities in the current–voltage, *I*(*V*), curves. In a separate control experiment, we observed that an irreversible decomposition of gold into separate clusters starts at 150–200 K.

Second, a submonolayer (~1%) of organic molecules was deposited on the electrodes by quench condensation. The sample was annealed at low temperature (below 70 K) allowing thermally activated motion of the organic molecules, while monitoring the nanogap conductance at a source–drain bias of 400 mV. When the conductance changed stepwise, indicating the trapping of a single molecule in the nanogap, the device was cooled to a temperature of 4.2 K where all transport measurements were taken. This entire process was repeated successfully for three independent devices.

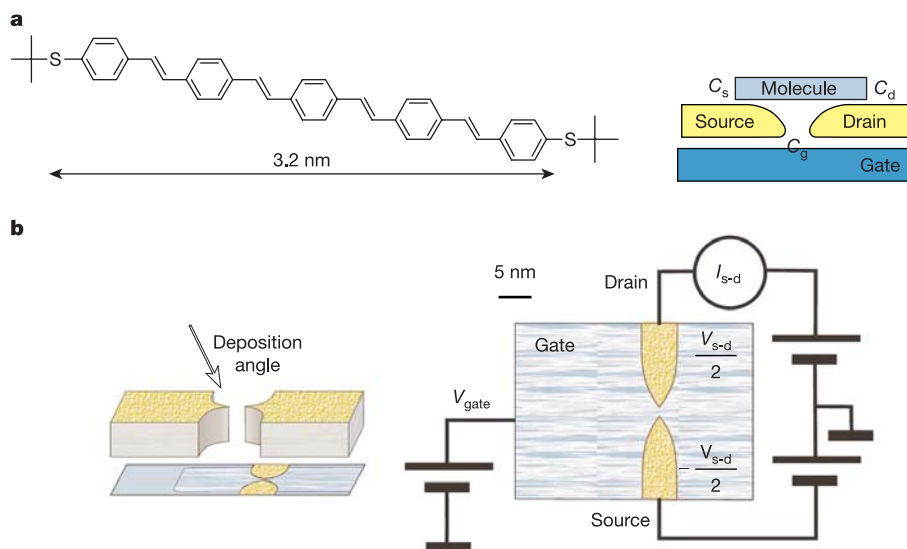


Figure 1 Device and experiment. **a**, Molecular structure of OPV5 and schematic experimental set-up. **b**, Schematic representation of the device preparation procedure.

By changing the gate voltage (V_g) of the SET with a single molecule in the nanogap in small steps from -4.3 V to $+4.3$ V while measuring the source–drain current–voltage characteristics (I_{s-d} – V_{s-d}) at each step (Fig. 2), we were able to probe eight different transmitting (open) states of the SET. This is summarized in Fig. 2a, where the differential conductance (dI_{s-d}/dV_{s-d}) of the SET is plotted in colour code as a function of the gate voltage (V_g) and source–drain voltage (V_{s-d}). The dark diamonds in Fig. 2a correspond to zero-current regions, where the low bias transport is blocked. Not all of the dark diamonds are complete, because the measurements were taken in a limited V_{s-d} range of ± 75 mV (at higher source–drain voltages the sample was unstable, possibly owing to the switching between different molecular conformations).

Two features emerge from the data set: (1) all diamond edges (the lines separating transmitting and closed regions) are straight and possess only two characteristic slopes, one positive and one negative; and (2) contrary to cases where SETs are dominated by Coulomb charging alone, the zero-bias open states are distributed in a non-equidistant way along the V_g axis, implying that effects other than simple Coulomb charging dominate the properties of

our device. Around $V_g \approx 0$ these states clearly appear in bunches of two.

The linearity of the diamond edges allows us to express the electrostatic interaction between the molecule and gate and lead electrodes in terms of three effective capacitances (C_s , C_d , C_g , Fig. 1a). This is known as the capacitance model^{15,16}. We note that linear diamond edges have also been observed in other experiments with gated molecules^{7,8}. The data in Fig. 2a show that the same set of capacitances characterizes the SET for all charged states of the molecule. This in turn proves that we are dealing with a system with just two tunnelling gaps—that is, with only one molecule connecting the source and drain electrodes—because a system with more than one quantum dot always has more than two characteristic slopes¹⁷.

To further investigate the second feature of our data, we use the capacitance model introduced above to extract the energies required to add an extra charge to a particular molecular ion. According to the theory of SET operation^{15,18}, the molecule can pass charge through the SET at every open state of the transistor by switching between two states with charge $(n+1)e$ and ne , where e denotes the electron charge (-1.6×10^{-19} C). The corresponding open state

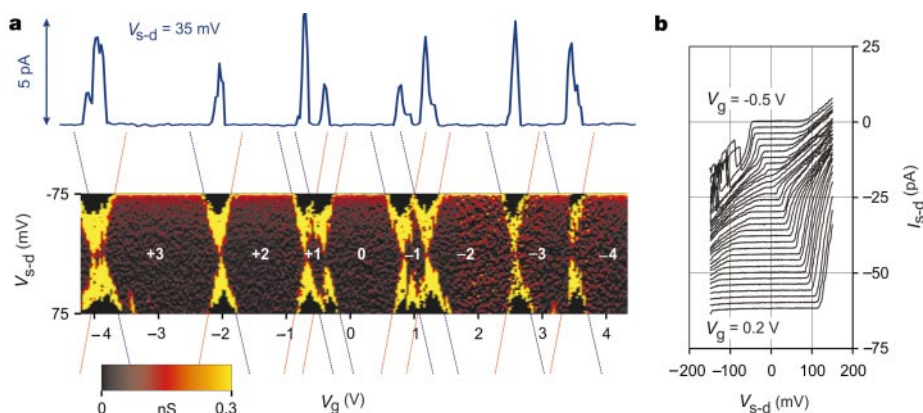


Figure 2 Experimental results. **a**, Measurements of the differential conductance (dI_{s-d}/dV_{s-d}) as function of V_{s-d} and V_g . All red lines, and all blue lines, have identical slopes, as discussed in the text. The full solid line at the top of the figure shows a

representative I_{s-d} – V_g trace. **b**, Examples of current–voltage curves I_{s-d} –(V_{s-d}) for a single OPV5 molecule obtained at different gate potentials V_g (temperature $T = 4.2$ K). Curves are shifted vertically for clarity.

gate voltages $V_g(n)$ are defined by matching the electrochemical potentials of the electrodes and the OPV molecule: $\mu_{\text{ELECTRODE}} = \mu_{\text{OPV}}(n, V_g(n))$. The electrochemical potential $\mu_{\text{OPV}}(n, V_g)$ is the difference between the ground state energies E for the two differently charged ions at a gate voltage V_g : $\mu_{\text{OPV}}(n, V_g) = E(n, V_g) - E(n-1, V_g)$. According to the capacitance model, the gate voltage shifts the electrochemical potential of the molecule linearly, allowing us to relate the diamond sizes along the gate axis to the molecular charging (electron addition) energies, $I(n \rightarrow n+1)$, by:

$$I(n \rightarrow n+1) = \mu_{\text{OPV}}(n+1, 0) - \mu_{\text{OPV}}(n, 0) \\ = (C_g / (C_s + C_d + C_g)) e (V_g(n) - V_g(n+1)) \quad (1)$$

where the scaling factor $C_g / (C_s + C_d + C_g)$ equals the ratio of the Coulomb diamond size along bias and gate axes multiplied by 2 as also shown in ref. 16 for a metallic island. We can use any of the two complete dark diamonds in Fig. 2a to estimate $C_g / (C_s + C_d + C_g)$, which turns out to be ~ 0.2 .

In terms of atomic energies the addition energy is the difference between the ionization energy and the electron affinity¹⁸. For a quantum dot treated within a single-particle picture, the addition energy I equals the energy difference between two adjacent states plus the Coulomb energy $e^2 / (C_s + C_d + C_g)$ (ref. 18). For the neutral molecule, this can be expressed as:

$$I(0 \rightarrow 1) = \varepsilon^{\text{LUMO}} - \varepsilon^{\text{HOMO}} + e^2 / (C_s + C_d + C_g) \quad (2)$$

where $\varepsilon^{\text{LUMO}} - \varepsilon^{\text{HOMO}}$ is the single-particle gap between the highest occupied molecular orbital (HOMO) and the lowest unoccupied molecular orbital (LUMO).

From Fig. 2 and the addition energies listed in Table 1, it is clear that the addition or removal of an electron from the molecule in a particular electronic configuration is not reduced to a simple change of electrostatic energy in the system, but also involves the energy related to the occupation of quantum-mechanical states. In metallic islands or nanotubes, such effects are small owing to their high density of states¹⁹. For a molecule, we deal with the opposite

situation where the discretization of electronic levels should dominate.

The data in Fig. 2a and Table 1 clearly reveal a symmetric bunching effect around $V_g = 0$, allowing us to use a symmetry argument for assigning this region to the neutral molecule. On the basis of this assignment, we have tabulated the charge on the molecule (redox state) in Table 1, together with independent molecular addition energies provided by the SET for both positively and negatively charged molecules. For comparison, we also list addition energies obtained from electrochemistry²⁰ and quantum-chemical calculations (Methods) for the negatively and positively charged redox states, respectively. The electrochemical and computational values have been scaled by the factors given in Table 1 to best match the trends in the SET data. Furthermore, the addition energy

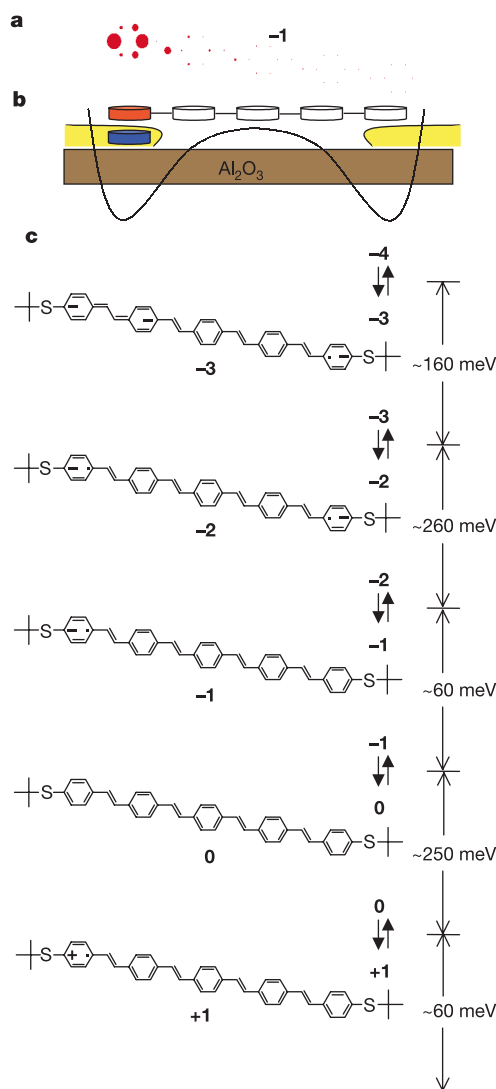


Figure 3 Calculations of charge confined on the OPV5 molecule, and valence bond models of the redox states involved. **a**, Charge distribution calculated by a Hubbard model for a single charge trapped on the OPV5 molecule and stabilized by its image charge; the size of the red circles is proportional to the atomic charge density. **b**, Schematic representation of the effective potential of a charge confined on the OPV5 molecule and influenced by its image charges in the source and drain electrodes (full line). The blue and red-coloured rings illustrate the equilibrium position of the image charge in the electrode and the charge on the molecule, respectively. **c**, Selected valence bond models of the involved redox states (left side) and corresponding addition energies (right side) between open states involving the redox couples indicated by the double arrows.

Table 1 Addition energies and redox states

Redox state	n	$V_g(n)$ (V)	$I(n \rightarrow n+1)^{\text{SET}}$ (meV)			$I(n \rightarrow n+1)^{\text{other}}$ (meV)	
-4	4	3.53*	†	‡	*		
-3	3	2.64			162	60§	
-2	2	1.20			262	280§	
-1	1	0.84			60	66	70§
0	0	-0.42			312	230	2,500
1	-1	-0.77	70	56	63	39¶	66#
2	-2	-2.13	390	266	247	350¶	
3	-3	-4.10			358	165¶	

Symbols are defined in the text.

*Data in these two columns are from Fig. 2a, representing the essence of many measurements on the same device/molecule.

†‡Data in these two columns are from two additional devices prepared independently. The ratio $C_g / (C_s + C_d + C_g) = 0.2$.

§From electrochemical midway potentials (ref. 20) multiplied by a scaling factor of 0.5.

|| Estimated from absorption edge of OPV5 in solution situated at 450 nm (own data).

¶ Estimated from computed ionization potentials (AM1, dielectric constant = 2.5, as suggested from data in ref. 25) multiplied by a scaling factor of 0.16.

Addition energies calculated by a Hubbard model (see Methods) representing the molecule as 38 p_z carbon orbitals with an on-site repulsion U and a transfer integral t between connected atoms. Image charges have been included by changing the on-site energy on one or both terminal benzene ring(s) by an amount equal to the mirror image charge energy, Φ , given by $\Phi_{\text{image}} = 27.2 \text{ (eV\AA)} / (\varepsilon_r a)$, where ε_r is the relative dielectric constant and a the distance between the charge and its image in angstroms. The values listed in Table 1 are obtained by fitting t to the optical HOMO–LUMO gap of the pristine OPV5 molecule (2,500 meV). The two independent parameters (ε_r , a) and U are obtained by fitting the model to the experimental addition energies obtained for $n = -1, 0, +1$. The resulting values are: $t = 3.92 \text{ eV}$, $U = 4.27 \text{ eV}$, $(\varepsilon_r, a) = 4.7 \text{ \AA}$. The value of (ε_r, a) is in good agreement with the expected distance between the plane of the benzene ring and its image when the benzene ring is in van der Waals contact with the electrode. The charges are thus expected to be separated by roughly twice the van der Waals radius of sp^2 -hybridized carbon ($\sim 3 \text{ \AA}$). U is also well within the expected range for on-site repulsion in organic molecules.

for the neutral molecule ($\sim \varepsilon^{\text{LUMO}} - \varepsilon^{\text{HOMO}}$), which is not directly available from electrochemical measurements, has been estimated from the onset of the absorption spectrum of OPV5.

The comparison between SET addition energies and the electrochemical and computational values shows similar evolutions for both the positive and negative redox states. But for the neutral molecule, there occur marked deviations among the data; the spectroscopic HOMO–LUMO gap (~ 2.5 eV) is one order of magnitude larger than the HOMO–LUMO gap extracted from the SET data, according to equation (2) (~ 0.2 eV). The rather large scaling factors required to match the electrochemical and computational values to the SET data and the dramatic change in the HOMO–LUMO gap strongly indicate that the intrinsic electronic level spacings of the molecule have been significantly altered in the metallic junction. On the basis of simple model calculations listed in Table 1, we suggest that image charges generated in the source and drain electrodes by the charges on the molecule are the most likely origins of this effect. The calculations are carried out using a Hubbard model of the molecule that includes image charges (Fig. 3a, b, Table 1). The results confirm the intuitive physical picture that a charge trapped on the molecule is strongly attracted to its image in the electrode, leading to a localization of the charge close to the electrode. Simple application of Coulomb's law then suggests that the image charge can change the local potential on the benzene ring in contact with the electrode by several eV. The resulting addition energies calculated for $n = \{-1, 0, 1\}$ are in good agreement with the SET experiment for very realistic values of the fitting parameters in the model (see the last footnote in Table 1 for details). These theoretical results are further corroborated by recent spectroscopic studies²¹ showing that the LUMO of hexafluorobenzene (C_6F_6) moves by 0.7 eV towards the middle of the HOMO–LUMO gap when this molecule is deposited on a metal surface and ionized through a charge transfer reaction. All together, these findings suggest that the initially injected charges have a localized character on the molecular backbone and can be described as confined polarons. This unorthodox localized charge representation provides a straightforward explanation for the bunching of the molecular charging energies around the neutral state; the energy difference between the bunched states simply represents the rather small energy required to charge both ends of the molecule in two successive steps (Fig. 3c). Further charging of the molecule leads to increased repulsion among the charges, which translates into larger charging energies, as observed experimentally.

We have reported SET data on a single conjugated molecule taken through a series of distinct redox states. Comparison between the intrinsic (redox) addition energies and those measured in the SET reveals a strong perturbation of the molecular properties that we attribute to image charges in the metal electrodes. The results show that the molecule cannot be conceptually separated from the electrodes owing to these electrostatic image charge effects, even when the electronic overlap between the electrodes and the molecule is weak, as evidenced by the observation of single-electron tunnelling phenomena. □

Methods

Total energies from computed ionization potentials

Total energies were calculated at the semiempirical Hartree–Fock Austin Model 1 (AM1) level²² coupled to a full configuration interaction (CI) scheme within an active space (encompassing five frontier electronic levels) to account for electron correlation effects (see penultimate footnote in Table 1). The wavefunctions have been treated with the restricted open-shell Hartree–Fock (ROHF) formalism for systems with an odd number of electrons. The choice of the AM1 method is driven by its good track record in reproducing the heat of formation of organic molecules. The electrostatic screening of the charges injected into the conjugated backbone by the surrounding medium also has to be taken into account to evaluate properly the total energies of the various redox states. This was achieved in our calculations by using the conductor-like screening model (COSMO)²³ implemented in the Ampac package (Ampac 6-55 ed.; Semichem). The self-consistent reaction field (SCRF) model treats the solvent as an uniform polarizable medium, with a dielectric constant ε , in which the solute is positioned in a suitably shaped cavity. This

model is known to provide realistic estimates of molecular total energies in solution when combined to the AM1 parameterization²⁴.

Addition energies from a Hubbard model

The addition energies of redox states at zero source–drain bias are calculated using the formula $I(n \rightarrow n+1) = E_{\text{tot}}(n+1) + E_{\text{tot}}(n-1) - 2E_{\text{tot}}(n)$, where $E_{\text{tot}}(n)$ is the total energy of molecule with n electrons (see last footnote in Table 1). Included in this energy is the sum of energies of occupied one-electron states as calculated in the Hartree–Fock approximation of the Hubbard model, corrected for double counting of Coulomb repulsion among electrons and the Coulomb energy of the ions experiencing the potential from the mirror charges in the gates. We note that the charge distribution of the molecule in the different redox states are not simple scalings of a fixed charge distribution, as is being assumed in the orthodox capacitance model. Work is in progress (P.H. and K. Flensburg) on giving a full account of the generalization of the capacitance model suggested by the experimental findings of the present work.

Received 23 June; accepted 13 August 2003; doi:10.1038/nature02010.

1. Averin, D. V. & Likharev, K. K. *Mesoscopic phenomena in Solids* (eds Al'tshuler, B. L., Lee, B. A. & Webb, R. A.) 173–271 (North-Holland, New York, 1991).
2. Kouwenhoven, L. P. et al. in *Mesoscopic Electron Transport* (eds Sohn, L. L., Kouwenhoven, L. P. & Schön, G.) 105, Ser. E, Vol. 345 (Kluwer, Dordrecht, 1997).
3. Ralph, D. C., Black, C. T. & Tinkham, M. Gate-voltage studies of discrete electronic states in aluminum nanoparticles. *Phys. Rev. Lett.* **78**, 4087–4090 (1997).
4. Klein, D. L., Roth, R., Lim, A. K. L., Alivisatos, A. P. & McEuen, P. L. A single electron transistor made of a cadmium selenide nanocrystal. *Nature* **389**, 699–701 (1997).
5. Cobden, D. H. et al. Transport spectroscopy of single-walled carbon nanotubes. *Physica B*, **249–251**, 132–135 (1998).
6. Tans, S. J., Devoret, M. H., Groeneveld, R. J. A. & Dekker, C. Electron-electron correlations in carbon nanotubes. *Nature* **394**, 761–764 (1998).
7. Park, H. et al. Nanomechanical oscillations in a single- C_{60} transistor. *Nature* **407**, 57–60 (2000).
8. Park, J. et al. Coulomb blockade and the Kondo effect in single-atom transistors. *Nature* **417**, 722–724 (2002).
9. Liang, W., Shores, M. P., Bockrath, M., Long, J. R. & Park, H. Kondo resonance in a single-molecule transistor. *Nature* **417**, 725–729 (2002).
10. Nygård, J., Cobden, D. H. & Lindelof, P. E. Kondo physics in carbon nanotubes. *Nature* **408**, 342–346 (2000).
11. Zhitenev, N. H., Meng, H. & Bao, Z. Conductance of small molecular junctions. *Phys. Rev. Lett.* **88**, 226801 (2002).
12. Stühr-Hansen, N., Christensen, J. B., Harrit, N. & Bjørnholm, T. Novel synthesis of protected thiol end-capped stilbenes and oligo(phenylenevinylene)s (OPVs). *J. Org. Chem.* **68**, 1275–1282 (2003).
13. Kubatkin, S. E., Danilov, A. V., Olin, H. & Claeson, T. Tunneling through a single quench-condensed cluster. *J. Low-Temp. Phys.* **118**, 307–316 (2000).
14. Park, H., Lim, A. K. L., Alivisatos, A. P., Park, J. & McEuen, P. L. Fabrication of metallic electrodes with nanometer separation by electromigration. *Appl. Phys. Lett.* **75**, 301–303 (1999).
15. van Houten, H., Beenakker, C. W. J. & Staring, A. M. in *Single Charge Tunnelling and Coulomb Blockade Phenomena in Nanostructures* Ch. 5 (Plenum, New York, 1992).
16. Likharev, K. K. Single electron devices and their application. *Proc. IEEE* **87**, 606–632 (1999).
17. Danilov, A. V., Golubev, D. S. & Kubatkin, S. E. Tunneling through a multigrain system: Deducing sample topology from nonlinear conductance. *Phys. Rev. B* **65**, 125312–125320 (2002).
18. Kouwenhoven, L. P., Austing, D. G. & Tarucha, S. Few-electron quantum dots. *Rep. Prog. Phys.* **64**, 701–706 (2001).
19. Cobden, D. H. & Nygård, J. Shell filling in closed single-wall carbon nanotube quantum dots. *Phys. Rev. Lett.* **89**, 46803 (2002).
20. Heinze, J., Mortensen, J., Müllen, K. & Schenk, R. The charge storage mechanism of conducting polymers: A voltammetric study on defined soluble oligomers of the phenylene-vinylene type. *J. Chem. Soc. Chem. Commun.* 701–703 (1987).
21. Vondrak, T. & Zhu, X.-Y. Two-photon photoemission study of heterogeneous electron transfer: C_6F_6 on Cu(111). *J. Phys. Chem. B* **103**, 3449–3456 (1999).
22. Dewar, M. J. S., Zebisch, E. G., Healy, E. F. & Stewart, J. J. P. A new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc.* **107**, 3902–3909 (1985).
23. Klamt, A. & Schürmann, G. J. Cosmo—a new approach to dielectric screening in solvents with explicit expressions for the screening energy and its gradient. *J. Chem. Soc. Perkin Trans. 2*, 5, 799–805 (1993).
24. Pourtois, G., Beljonne, D., Cornil, J., Ratner, M. A. & Brédas, J. L. Photoinduced electron-transfer processes along molecular wires based on phenylenevinylene oligomers: A quantum-chemical insight. *J. Am. Chem. Soc.* **124**, 4436–4447 (2002).
25. Comoretto, D. et al. Optical constants of highly stretch-oriented poly(p-phenylene-vinylene): A joint experimental and theoretical study. *Phys. Rev. B* **62**, 10173–10184 (2000).

Acknowledgements We thank K. Flensburg for discussions. Financial support from the European Union under the IST programme 'NANOMOL' (initiated by M. Persson) is acknowledged. Work in Denmark is also supported by the Danish Research Council. The work in Arizona is supported by the Office of Naval Research, National Science Foundation, and the IBM Shared University Research Program. The work in Mons is supported by the Belgian Federal Government 'InterUniversity Attraction Pole in Supramolecular Chemistry and Catalysis' and the Belgian National Fund for Scientific Research. J.C. is a research fellow of the FNRS. A.D. was supported by the Swedish SSE.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to T.B. (tb@nano.ku.dk).

Zero thermal expansion in YbGaGe due to an electronic valence transition

James R. Salvador¹, Fu Guo², Tim Hogan² & Mercouri G. Kanatzidis¹

¹Department of Chemistry and Centre for Fundamental Materials Research, and

²Department of Electrical and Computer Engineering, Michigan State University, East Lansing, Michigan 48824, USA

Most materials expand upon heating. Although rare, some materials expand on cooling, and are said to exhibit negative thermal expansion (NTE); but the property is exhibited in only one crystallographic direction. Such materials include silicon and germanium¹ at very low temperature (<100 K) and, at room temperature, glasses in the titania–silica family², Kevlar, carbon fibres, anisotropic Invar Fe–Ni alloys³, ZrW₂O₃ (ref. 4) and certain molecular networks⁵. NTE materials can be combined with materials demonstrating a positive thermal expansion coefficient to fabricate composites exhibiting an overall zero thermal expansion (ZTE). ZTE materials are useful because they do not undergo thermal shock on rapid heating or cooling. The need for such composites could be avoided if ZTE materials were available in a pure form. Here we show that an electrically conductive intermetallic compound, YbGaGe, can exhibit nearly ZTE—that is, negligible volume change between 100 and 400 K. We suggest that this response is due to a temperature-induced valence transition in the Yb atoms. ZTE materials are desirable to prevent or reduce resulting strain or internal stresses in systems subject to large temperature fluctuations, such as in space applications and thermomechanical actuators.

YbGaGe and YbGaSn crystallize in the hexagonal *P6₃/mmc* space group (MoC₂ according to Pearson's structure type)⁶ (see ref. 7 and references therein). The structure is a derivative of the MgB₂

structure type and is characterized by puckered boron nitride-like GaGe and GaSn layers (for example, A, B, C and D in Fig. 1a). In YbGaGe, the Ga and Ge positions could not be distinguished from the X-ray data and the assignment was made on the basis of expected bond distances around these atoms and by analogy to the assignment of YbGaSn. The bond distances within the layers for Ga–Ge and Ga–Sn are 2.520(1) Å and 2.696(1) Å respectively, indicating strong bonding.

There are four such layers in the unit cell, ordering in an ABCD fashion. The AB and CD layers form respective pairs sandwiching the Yb(1) atoms. Within each pair—AB and CD—the Ga atoms are in alignment (that is, they face each other) so that they present the closest approach between layers in each pair and provide the binding site for the Yb(1) atoms. The Yb(2) atoms are found between the AB and CD pairs and bind to Ge or Sn atoms (Fig. 1a). Yb(1) has a trigonal prismatic coordination of Ga atoms (Fig. 1c), with bond distances of 2.921(2) Å in YbGaGe and 2.966(2) Å in YbGaSn.

This anomalously large difference between the Yb(1)–Ga bond distances in the two compounds was a first indication of the unusual nature of YbGaGe. Yb(2) is coordinated octahedrally by Ge (or Sn) atoms with distances of 3.078(1) Å and 3.238(1) Å respectively (Fig. 1d). Valence bond sum calculations, carried out to assess the oxidation state of the Yb ions in YbGaSn and YbGaGe, showed that both ions in YbGaSn were in the +2 oxidation state; in YbGaGe, however, Yb(1) was calculated to be +2.6 and Yb(2) was +2.0, suggesting an intermediate valence compound.

Another peculiar characteristic of these two structures that led us to investigate their properties further was the difference in the Ga–Ga distances between adjacent AB and CD layers that surround Yb(1). In YbGaSn this distance is 2.994(1) Å, and is considered short enough for a weak interaction. Despite the presence of the smaller Ge atom in YbGaGe, the Ga–Ga distance is much longer at 3.248(1) Å, and cannot be considered a bonding interaction. This lengthening is the consequence of electron transfer from a formally Yb²⁺ ion to a Ga–Ga bond, resulting in partial oxidation to Yb³⁺ and reductive cleavage of the Ga–Ga bond. The shortened Ga–Yb(1)

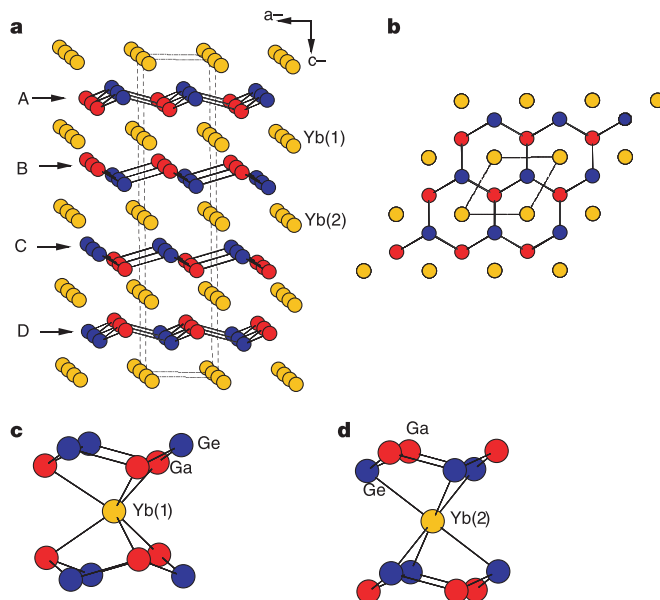


Figure 1 Crystal structure details of YbGaGe. **a**, Unit cell of YbGaGe. Red atoms are Ga, blue atoms are Ge and yellow atoms are Yb. The layers A, B, C and D are indicated. **b**, Structure viewed down the *c* axis of the *P6₃/mmc* unit cell. **c**, Immediate trigonal prismatic coordination environment of Yb(1). **d**, Octahedral coordination of Yb(2). Atomic coordinates ($\times 10^4$) and equivalent isotropic displacement parameters ($\text{\AA}^2 \times 10^3$) for

YbGaGe/YbGaSn are as follows, where U_{eq} is defined as one-third of the trace of the orthogonalized U_{ij} tensor. For Yb(1), (x, y, z) = (0, 0, 2.500) and U_{eq} = 14(1)/16(1). For Yb(2) (x, y, z) = (0, 0, 0) and U_{eq} = 10(1)/11(1). For Ge/Sn, (x, y, z) = (3,333, -3,333, 6,128(1)/6,146(1)) and U_{eq} = 9(1)/12(1). For Ga, (x, y, z) = (3,333, -3,333, 1,531(2)/1,633(2)) and U_{eq} = 14(1)/14(1).

distance and compressed c axis in the Ge analogue is due to the fact that Yb(1) in YbGaGe has mixed valency, $2+3+$, and thus has a smaller radius.

There is a certain composition width associated with the phase of

YbGaGe. For example, the phase can tolerate a certain deviation from the ideal stoichiometry in the form of $\text{YbGa}_{1+x}\text{Ge}_{1-x}$ (when the material is made from Ga-rich conditions) and $\text{YbGa}_{1-x}\text{Ge}_{1+x}$ (when the material is made from Ge-rich conditions) and this is evident in

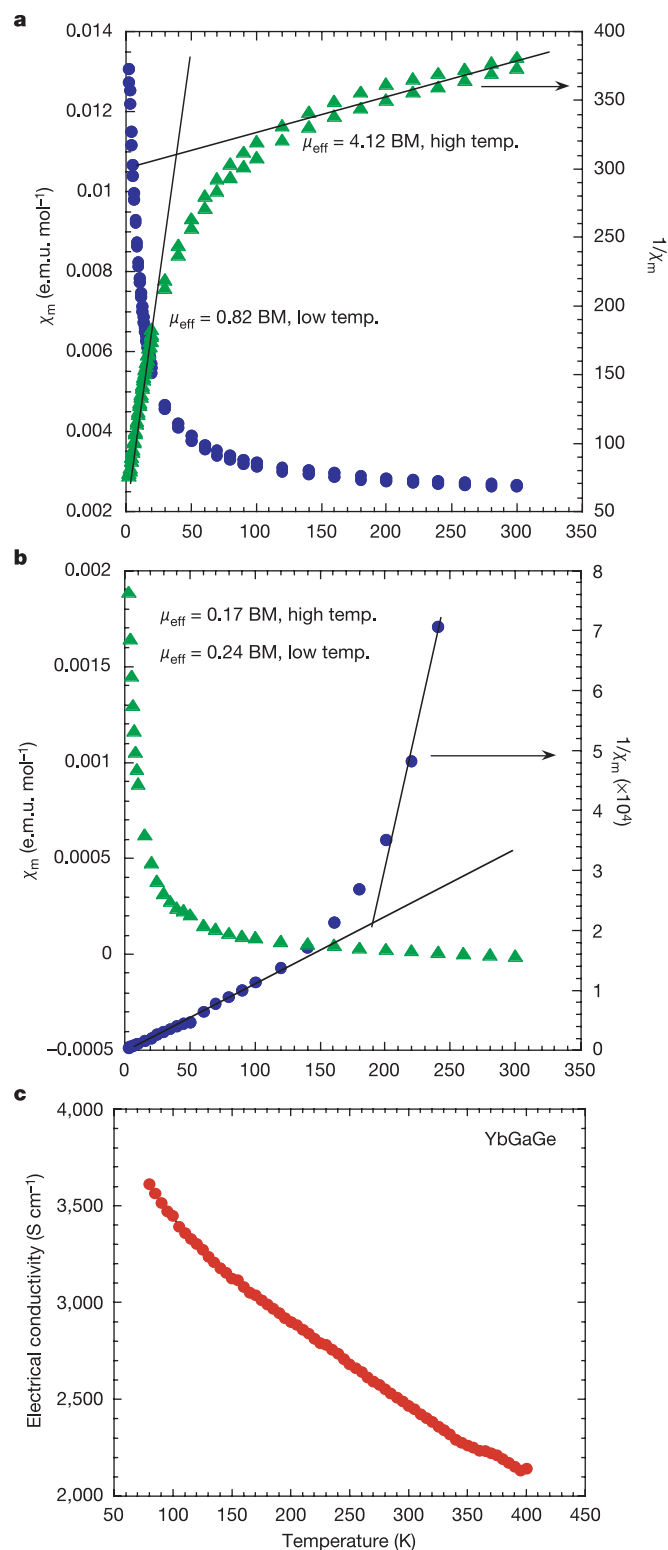


Figure 2 Magnetic susceptibility and electrical conductivity data. **a**, YbGaGe and **b**, YbGaSn (both χ_m and $1/\chi_m$). Applied field was 200 Oe for YbGaGe and 500 Oe for YbGaSn. **c**, Variable-temperature electrical conductivity for YbGaGe. Transport measurements were made using the standard four-probe configuration.

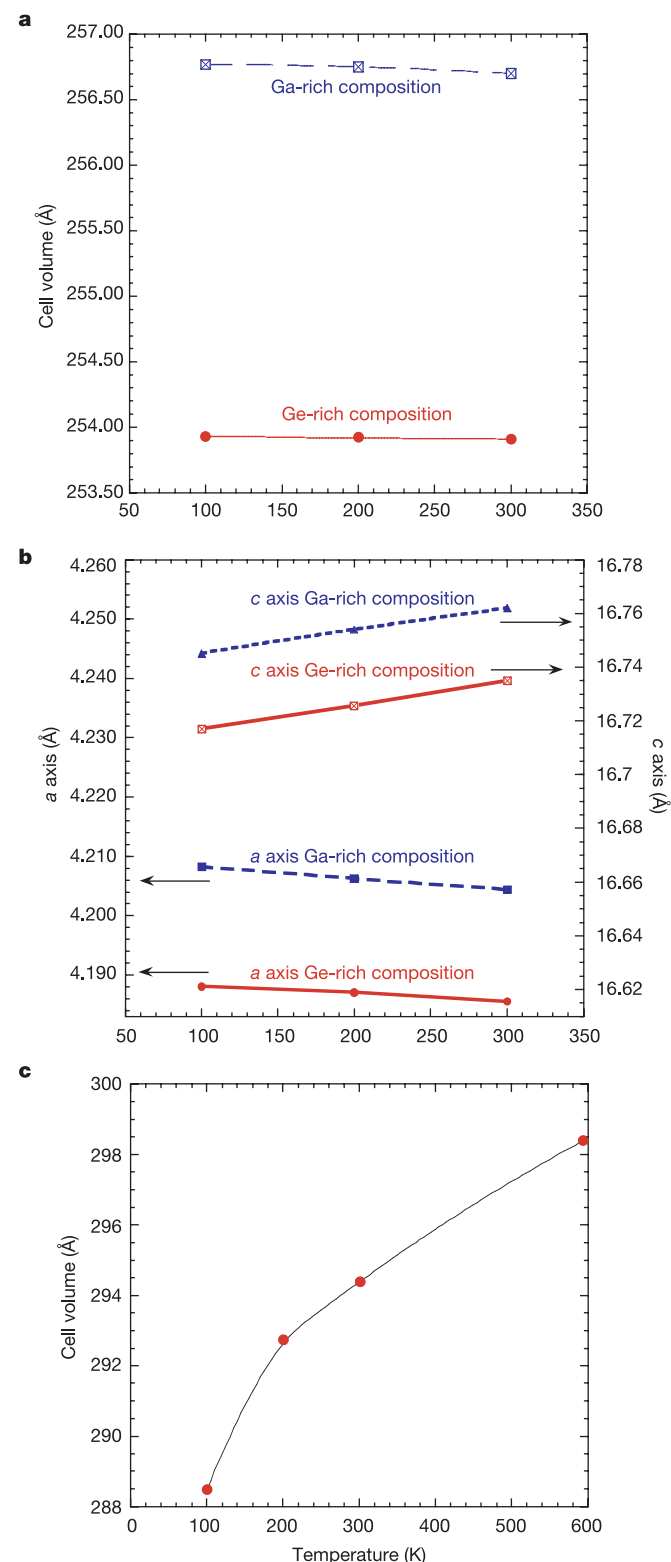


Figure 3 Thermal expansion data. **a**, Unit-cell volume versus temperature for YbGaGe. **b**, a and c cell edge lengths versus temperature for YbGaGe. **c**, Unit-cell volume versus temperature for YbGaSn. Single-crystal measurements were made by collecting a full set of data at room temperature, 200 K and 100 K. The lowest attainable temperature was 100 K. All data sets were collected on the same crystal.

the slightly different cell parameters of the crystals (for example, at 200 K, $a = 4.2064(7)\text{\AA}$, $c = 16.758(5)\text{\AA}$ for $\text{YbGa}_{1+x}\text{Ge}_{1-x}$ versus $a = 4.1867(3)\text{\AA}$, $c = 16.7275(22)\text{\AA}$ for $\text{YbGa}_{1-x}\text{Ge}_{1+x}$).

Crystallographic data were collected at 200 and 100 K for $\text{YbGa}_{1-x}\text{Ge}_{1+x}$ to investigate the structural changes taking place upon cooling. At 200 K, the Yb(1)–Ga bonds and the Yb(1)–Ge bonds are 2.8715(32) and 3.797(3) Å respectively and the Ga–Ge distance is 2.523(2) Å. Upon cooling to 100 K, the Yb(2)–Ge bonds both elongate to 3.0825(16) Å, while the Yb(1)–Ga is unchanged at 2.869(3) Å. The Ga–Ge distance also remains constant at 2.523(3) Å. The composition $\text{YbGa}_{1+x}\text{Ge}_{1-x}$ was measured between 200 and 300 K and it was found that the bond distances stayed constant over the temperature range, with Yb(1)–Ga distances of 2.9182(15) Å and 2.9178(15) Å at 200 and 300 K respectively, Yb(2)–Ge distances of 3.0803(16) Å and 3.0803(15) Å for 200 and 300 K respectively and Ga–Ge distances of 2.5211 Å and 2.5199(12) Å at 200 and 300 K.

Clearly, this suggests interesting behaviour in thermal expansion, which depends on the exact composition of $\text{YbGa}_{1+x}\text{Ge}_{1-x}$. We have observed overall zero volumetric thermal expansion in most samples coming from Ga-rich and Ge-rich synthetic conditions, with one showing NTE.

The critical clue for the origin of anomalous thermal expansion behaviour in YbGaGe came from the magnetic susceptibility data, which showed a temperature-induced variation in the magnetic moment of Yb atoms in this system and consequently a change in the $\text{Yb}^{2+/3+}$ ratio. The data showed an increase in the Yb^{2+} population at low temperature, indicating a volume expansion associated with the Yb^{3+} to Yb^{2+} transition. In contrast, YbGaSn is only weakly paramagnetic, with no change in the Yb^{2+} fraction (Yb^{3+} is a paramagnetic ion with calculated free ion effective magnetic moment of 4.54 BM, whereas Yb^{2+} is diamagnetic). Figure 2a shows the very different behaviour of the molar susceptibility as a function of temperature for YbGaSn and YbGaGe. YbGaGe shows a high-temperature region from 150 to 300 K that has a large μ_{eff} of 4.12 μ_{B} , while below 30 K, μ_{eff} drops to only 0.82 μ_{B} . The region between 30 and 150 K has a continuously changing slope and cannot be fitted to the Curie–Weiss law. This is consistent with a progressive change in Yb valence.

The magnetic data and structural data clearly show that when YbGaGe is cooled electrons move to the Yb^{3+} ions (f^{13} to f^{14} reduction) by depopulating the conduction band, thereby increasing the fraction of Yb^{2+} ions. Since Yb^{2+} is much larger than Yb^{3+} (1.16 versus 1.008 Å radius), an overall increase in the total volume is observed (see below). By comparison, YbGaSn shows a very small effective moment μ_{eff} of 0.17 μ_{B} from 100 to 300 K and 0.24 μ_{B} below 90 K (Fig. 2b). These values are consistent with predominantly Yb^{2+} in the tin analogue.

To observe the anomalous behaviour in the unit cell parameters and cell volume upon cooling, we performed X-ray diffraction experiments as a function of temperature for single crystals of $\text{YbGa}_{1+x}\text{Ge}_{1-x}$ and $\text{YbGa}_{1-x}\text{Ge}_{1+x}$ (Fig. 3a, b). Two of the unit cell constants (a , b) increase with falling temperature as expected for an expansion over the a – b plane. The c axis contracts by such an amount that the overall volume change is nearly zero (Fig. 3a). The magnitude of the NTE coefficient along the a direction was determined from the linear regression analysis of the temperature-dependent volume data to be $-1.3 \times 10^{-5} \text{ K}^{-1}$ for the Ge-rich compositions and -1.8×10^{-5} for the Ga-rich compositions^{3,8,9}. In contrast, YbGaSn shows a positive thermal expansion coefficient of $1.0 \times 10^{-4} \text{ K}^{-1}$ (Fig. 3c).

The mechanism responsible for the NTE behaviour of ZrW_2O_8 and related oxides is geometrical in origin and has to do with cooperative rotation of the oxide tetrahedra. YbGaGe does not undergo any type of structural rotation upon cooling, as the crystal refinements at the various temperatures reveal. Instead, the NTE in YbGaGe (along the a – b plane) and the overall ZTE derive from the internal electronic charge transfer of the Yb ions in the compound.

We note that valence fluctuations were claimed as the cause for NTE coefficients, for the quasi-one-dimensional platinum compounds $[\text{Pt}(\text{NH}_3)_4][\text{PtX}_2(\text{NH}_3)_4](\text{HSO}_4)_4$ (where $X = \text{Cl, Br, I}$)¹⁰.

This is a previously unknown mechanism of inducing NTE (or ZTE). One other example is $\text{Sm}_{0.9}\text{La}_{0.1}\text{S}$, which also shows an NTE coefficient of $1.4 \times 10^{-5} \text{ K}^{-1}$, but only at 20 K and requires 1.7 GPa of pressure to show the effect¹¹. Similarly, the compounds CeInCu_2 and $\text{Ce}_x\text{La}_{1-x}\text{Al}_2$ show NTE as a result of Kondo lattice effects but it is limited to 2–10 K (refs 12, 13). The structurally identical YbGaSn has a positive thermal expansion coefficient, which lends further credence to the argument that the anomalous thermal expansion of YbGaGe is rooted in electronic causes and not in geometric configuration changes.

YbGaGe is a metallic material, with polycrystalline compactions of it showing room-temperature conductivity of $\sim 2,400 \text{ S cm}^{-1}$. As seen in Fig. 2c, the conductivity of YbGaGe at 77 K is approximately 3,700 S cm^{-1} and decreases to 2,200 S cm^{-1} by 400 K. By comparison, the conductivity of ZrW_2O_8 is $< 10^{-8} \text{ S cm}^{-1}$ at 500 K.

The very few materials that exhibit NTE (or ZTE) are insulators and can only be used as ceramics, insulators, or in optics. Thus, there is a need for new materials with similar thermal expansion behaviour over a range of temperatures but that are capable of conducting electric current. YbGaGe combines these characteristics and promises to enable new applications. Furthermore, the electronic, rather than geometric, mechanism operating in YbGaGe may possibly be exploited in the future to design other ZTE or NTE materials through the suitable selection of mixed-valency metal systems and surrounding elements. □

Methods

Synthesis of YbGaGe and YbGaSn

The pure phase of YbGaGe (which is stable in air) was obtained by directly combining the elements in their stoichiometric ratios and heating first to 850 °C for 96 h followed by quick cooling to room temperature ($\sim 23^\circ\text{C}$) over 12 h. YbGaSn was prepared similarly by heating to 700 °C. Elemental analysis performed on several crystals indicate a Yb:Ga:Ge (or Sn) ratio of 1:1:1, consistent with the X-ray crystallographic analysis. Differential thermal analysis performed on YbGaGe showed no melting or any phase change up to 1,000 °C. Samples with excess Ga or excess Ge can be prepared accordingly by adjusting the Yb:Ga:Ge ratio.

X-ray diffraction data

Single-crystal X-ray diffraction data for YbGaGe and YbGaSn were collected at 293 K on a Siemens SMART Platform CCD diffractometer. The structure refinements were done with the SHELXTL package of crystallographic programs. Calculations were carried out with Valence Bond Calculator version 2.00 (ref. 14). YbGa₂ was used as the standard reference for Yb–Ga bond distances.

For YbGaGe: hexagonal $P6_3/mmc$ (no.194), $a = 4.2056(8)\text{\AA}$, $c = 16.780(5)\text{\AA}$, $Z = 3$, $V = 257.3(1)\text{\AA}^3$, $d_{\text{calc}} = 6.112 \text{ g cm}^{-3}$, $\mu = 43.28 \text{ mm}^{-1}$, $2.43^\circ < \theta(\text{MoK}\alpha) < 26.85^\circ$, number of total reflections = 1,352; unique reflections = 141 [$R_{\text{int}} = 0.0434$]; final R indices for all data, $R_1 = 0.0279$, $wR_2 = 7.24$ for $I > 2\sigma$.

For YbGaSn: hexagonal $P6_3/mmc$ (no.194), $a = 4.4352(9)\text{\AA}$, $c = 17.291(5)\text{\AA}$, $Z = 3$, $V = 294.6(1)\text{\AA}^3$, $d_{\text{calc}} = 6.113 \text{ g cm}^{-3}$, $\mu = 36.471 \text{ mm}^{-1}$, $2.43^\circ < \theta(\text{MoK}\alpha) < 26.00^\circ$, total reflections = 1,959; unique reflections = 169 [$R_{\text{int}} = 0.046$]; final R indices for all data, $R_1 = 0.036$, $wR_2 = 0.092$ for $I > 2\sigma(I)$.

Received 5 March; accepted 19 August 2003; doi:10.1038/nature02011.

- Kagaya, H.-M. & Soma, T. Compression effect on specific heat and thermal expansion of Si and Ge. *Solid State Commun.* **85**, 617–621 (1993).
- Schultz, P. C. & Smyth, H. T. in *Amorphous Materials* (eds Douglas, R. & Ellis, B.) (Wiley, New York, 1970).
- Hausch, G., Bächer, R. & Hartmann, J. Influence of thermomechanical treatment on the expansion behavior of invar and superinvar. *Physica B* **161**, 22–24 (1989).
- Mary, T. A., Evans, J. S., Vogt, O. & Sleight, A. W. Negative thermal expansion from 0.3 to 1050 Kelvin in ZrW_2O_8 . *Science* **272**, 90–92 (1996).
- Baughman, R. H. & Galvao, D. S. Crystalline network with unusual predicted mechanical and thermal properties. *Nature* **365**, 735–737 (1993).
- Czybulka, A., Pinger, B. & Schuster, H.-U. New alkaline earth-gallium-silicides, germanides and stannides with AlB_2 type related structures. *Z. Anorg. Allg. Chem.* **579**, 151–157 (1989).
- Hoffman, R. D. & Pöttgen, R. AlB_2 -related intermetallic compounds—a comprehensive view based on group-subgroup relations. *Z. Kristallogr.* **216**, 127–145 (2001).
- Amos, T. G. & Sleight, A. W. Negative thermal expansion in orthorhombic NbOPO_4 . *J. Solid State Chem.* **160**, 230–232 (2001).
- Attfield, M. P. & Sleight, A. W. Exceptional negative thermal expansion in AlPO_4 . *Chem. Mater.* **10**, 2013–2016 (1998).
- Tsuyoshi, M. Negative thermal expansion and valence fluctuation in quasi-one-dimensional platinum compounds. *Phys. Rev. B* **35**, 6051–6058 (1987).

11. Oomi, G., Kuwahara, R., Kagayama, T. & Jung, A. Pressure-induced volume anomaly of intermediate valence compound $\text{Sm}_{0.9}\text{La}_{0.1}\text{S}$. *J. Magn. Magn. Mater.* **226–230**, 1182–1183 (2001).
12. de Visser, A., Bakker, K. & Pierre, J. Anomalous negative thermal expansion of CeInCu_2 . *Physica B* **186–188**, 577–579 (1993).
13. Lang, M., Schefzyk, R., Steglich, F. & Grewe, N. Negative thermal expansion in the Kondo system $\text{CeLa}_{1-x}\text{Al}_2$. *J. Magn. Magn. Mater.* **63–64**, 79–81 (1987).
14. Hormillosa, C., Healy, S. & Stephen, T. *Valence Bond Calculator version 2.00* (I. D. Brown, Institute for Material Research, McMaster University, Hamilton, Ontario, Canada, 1993).

Acknowledgements Financial support from the Department of Energy is gratefully acknowledged.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.G.K. (kanatzid@cem.msu.edu).

Reduction of soil carbon formation by tropospheric ozone under increased carbon dioxide levels

Wendy M. Loya¹, Kurt S. Pregitzer¹, Noah J. Karberg², John S. King¹ & Christian P. Giardina²

¹School of Forest Resources and Environmental Science, Michigan Technological University, Houghton, Michigan 49931, USA

²USDA Forest Service, North Central Research Station, Houghton, Michigan 49931, USA

In the Northern Hemisphere, ozone levels in the troposphere have increased by 35 per cent over the past century¹, with detrimental impacts on forest^{2,3} and agricultural⁴ productivity, even when forest productivity has been stimulated by increased carbon dioxide levels⁵. In addition to reducing productivity, increased tropospheric ozone levels could alter terrestrial carbon cycling by lowering the quantity and quality of carbon inputs to soils. However, the influence of elevated ozone levels on soil carbon formation and decomposition are unknown. Here we examine the effects of elevated ozone levels on the formation rates of total and decay-resistant acid-insoluble soil carbon under conditions of elevated carbon dioxide levels in experimental aspen (*Populus tremuloides*) stands and mixed aspen–birch (*Betula papyrifera*) stands. With ambient concentrations of ozone and carbon dioxide both raised by 50 per cent, we find that the formation rates of total and acid-insoluble soil carbon are reduced by 50 per cent relative to the amounts entering the soil when the forests were exposed to increased carbon dioxide alone. Our results suggest that, in a world with elevated atmospheric carbon dioxide concentrations, global-scale reductions in plant productivity due to elevated ozone levels will also lower soil carbon formation rates significantly.

Table 1 Total carbon and $\delta^{13}\text{C}$ of soils and the acid-insoluble fraction

	Total soil carbon		Acid-insoluble fraction	
	Total C (g m^{-2})	$\delta^{13}\text{C}$ (‰)	Total C (g m^{-2})	$\delta^{13}\text{C}$ (‰)
Control	5,385 ± 241	−26.7 ± 0.11	2,287 ± 185	−27.7 ± 0.06
Elevated O_3	5,237 ± 72	−26.9 ± 0.06	2,671 ± 224	−27.7 ± 0.11
Elevated CO_2	5,683 ± 480	−28.3 ± 0.10	2,594 ± 230	−29.0 ± 0.12
Elevated $\text{O}_3 + \text{CO}_2$	5,114 ± 369	−27.6 ± 0.10	2,473 ± 97	−28.5 ± 0.10

Determined for mineral soil from 0 to 20 cm. Values are mean ± s.e.

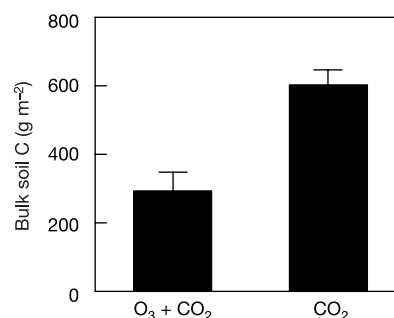


Figure 1 Total carbon incorporated into soils during 4 yr of exposure to elevated $\text{O}_3 + \text{CO}_2$ and elevated CO_2 . Values are mean ± 1 s.e.; $P < 0.01$.

Large areas of the Earth are exposed to concentrations of tropospheric ozone (O_3) that exceed levels known to be toxic to plants^{4,6}. In addition to reducing plant growth, exposure to elevated O_3 can also alter plant tissue chemistry⁷ and reduce allocation of carbon to roots and root exudates^{8–10}. Whereas the effects of O_3 on these aspects of plant biology have been widely investigated in chamber studies, examination of these effects on below-ground carbon cycling in intact forests only became possible in 1997 with the establishment of the FACTS-II (forest–atmosphere carbon transfer and storage) FACE (free-air carbon dioxide enrichment) experiment in Rhinelander, Wisconsin, USA. The long-term FACE experiment in Rhinelander examines how plant–plant and plant–microbe interactions may alter ecosystem responses to elevated O_3 and carbon dioxide (CO_2) through four treatments: control, elevated CO_2 , elevated O_3 and elevated $\text{O}_3 + \text{CO}_2$. In plots where O_3 and CO_2 are elevated, concentrations were maintained at ~150% of ambient levels¹¹. To examine the effects of atmospheric trace gases on both ecological interactions and on whole-ecosystem carbon cycling, each plot is split to include a pure aspen forest, a mixed aspen–birch forest and a mixed aspen–maple forest. These species were chosen because they are among the most widely distributed trees in northern temperate forests.

Here we compare soil carbon formation in aspen and aspen–birch subplots under elevated CO_2 and elevated $\text{O}_3 + \text{CO}_2$ (three plots each) to understand how exposure to O_3 under elevated CO_2 alters soil carbon formation. We used CO_2 derived from fossil fuel with its highly depleted ^{13}C signature to fumigate plant canopies in the elevated CO_2 and elevated $\text{O}_3 + \text{CO}_2$ plots. Leaf and root carbon inputs in the elevated CO_2 and elevated $\text{O}_3 + \text{CO}_2$ plots had a $\delta^{13}\text{C}$ signature of $-41.6 \pm 0.4\text{‰}$ (mean ± s.e.) in contrast to leaf and root inputs of $-27.6 \pm 0.3\text{‰}$ in the control plots. The $\delta^{13}\text{C}$ signature of soil carbon was $-26.7 \pm 0.2\text{‰}$ before fumigation and

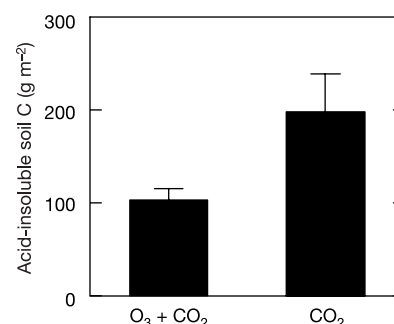


Figure 2 Carbon incorporated into the stable acid-insoluble fraction of soils during 4 yr of exposure to elevated $\text{O}_3 + \text{CO}_2$ and elevated CO_2 . Values are mean ± 1 s.e.; $P < 0.01$.

Table 2 Quantity and relative proportion of total soil carbon and soil carbon pools lost via microbial respiration in laboratory incubations

	Total carbon		Fumigation carbon		Pre-fumigation carbon	
	Cumulative loss	Relative mass loss	Cumulative loss	Relative mass loss	Cumulative loss	Relative mass loss
Control	1,372 ± 62	0.10 ± 0.01				
Elevated O ₃	1,392 ± 36	0.11 ± 0.00				
Elevated CO ₂	1,673 ± 85	0.12 ± 0.01	438 ± 23	0.27 ± 0.02	1235 ± 71	0.10 ± 0.00
Elevated O ₃ + CO ₂	1,342 ± 42	0.11 ± 0.01	267 ± 26	0.42 ± 0.06	1075 ± 36	0.09 ± 0.00
P	0.09	0.22	0.04	0.10	0.31	0.17

Cumulative loss (mg CO₂-C per kg soil). Relative mass loss (g CO₂-C per g soil-C) determined for day 281 of incubation. Values are mean ± s.e.

in control plots. Because plant carbon inputs to the soils fumigated with elevated CO₂ and elevated O₃ + CO₂ were depleted in ¹³C relative to soil carbon that existed before the experiment began, incorporation of new plant detritus into soil organic matter through time should decrease the δ¹³C of soil carbon. Using standard mixing models, we were able to determine the fraction of total soil carbon and acid-insoluble soil carbon derived from atmospheric CO₂ fixed by the trees over the 4 yr of experimental fumigation.

Elevated O₃ profoundly altered the ¹³C composition of soil carbon after 4 yr of fumigation. The less depleted δ¹³C signature of total soil carbon from the elevated O₃ + CO₂ treatment compared to the elevated CO₂ treatment (Table 1; *P* < 0.01) indicates that less carbon entered the soil when aspen and mixed aspen–birch forests were exposed to both elevated O₃ and CO₂. Carbon inputs to soil over 4 yr accounted for 10.7 ± 0.6% of the total soil carbon under elevated CO₂, but only 5.7 ± 0.9% under elevated O₃ + CO₂. Thus, elevated O₃ reduced total soil carbon formation by approximately 300 g C m⁻² compared to the amount formed under elevated CO₂ alone (Fig. 1).

Soils store carbon with both long and short turnover times, and to understand the carbon storage capacity of soils, it is necessary to partition soil carbon changes into decay-resistant and labile components. We used acid hydrolysis to isolate the most decay-resistant carbon in the soil¹². The acid-insoluble soil carbon fraction contains the longest-lived compounds, and is composed primarily of aromatic humic acids and residues of phenols, lignin and lignin-associated cellulose from newer plant residues¹³. However, after careful removal of plant residues before hydrolysis, new carbon inputs measured in this fraction are restricted to the most decomposition-resistant compounds. Compared to elevated CO₂ alone, simultaneous O₃ and CO₂ fumigation greatly reduced the quantity of carbon entering the acid-insoluble fraction in both aspen and aspen–birch soils, as indicated by the less depleted δ¹³C signature (Table 1; *P* < 0.04). Carbon incorporated into the soil since the initiation of the experiment accounted for approximately 9.1 ± 0.8% of the acid-insoluble carbon fraction in the elevated CO₂ plots, but only 4.2 ± 0.6% in the elevated O₃ + CO₂ plots. Thus, the elevated O₃ treatment reduced formation of acid-insoluble soil carbon by approximately 100 g C m⁻² (or 48%) compared with elevated CO₂ alone (Fig. 2).

Soil carbon formation is a balance between detrital inputs and decomposition losses. Observed reductions in soil carbon formation in the elevated O₃ + CO₂ treatment could result from reduced inputs, increased losses, or both. Previous research at this site suggests that plant detrital inputs are reduced under elevated O₃ + CO₂ relative to elevated CO₂ alone^{5,10}. To elucidate the effects of decomposition losses on soil carbon, we conducted a long-term laboratory incubation experiment. Despite greater microbial activity in soils from elevated CO₂ plots (Table 2), there was a strong trend of greater relative mass loss of carbon derived from fumigation in the elevated O₃ + CO₂ plots compared to the elevated CO₂ plots (Table 2; *P* < 0.1). This result suggests that the soil microbes responsible for carbon decomposition under elevated O₃ + CO₂ used a greater proportion (35%) of carbon formed since

fumigation began, leaving less of this carbon in the soil and in the acid-insoluble fraction. Possible reasons for greater consumption of recent carbon inputs include changes in microbial community composition¹⁴, microbial activity¹⁵, or carbon availability⁹. In turn, these below-ground changes could be driven by altered quantity and quality of detrital inputs^{5,10}.

The atmospheric trace gases CO₂ and O₃ are known to influence photosynthesis and ecosystem productivity in opposite ways. It is now clear that their interaction has the potential to cascade through ecosystems, altering soil carbon cycling. In response to reduced detrital carbon supply^{5,10} and increased microbial utilization, forest ecosystems exposed to both elevated O₃ and CO₂ accrued 51% less total, and 48% less acid-insoluble, soil carbon compared to ecosystems exposed only to elevated CO₂. Our results suggest that changes in the formation of soil carbon induced by tropospheric O₃ should be explicitly considered in models of soil carbon cycling under scenarios of elevated CO₂. □

Methods

Research area and field experiment

The FACTS-II FACE experiment in Rhinelander, Wisconsin, USA (49° 40.5' N, 89° 37.5' E, 490 m elevation) includes three experimental treatments (elevated CO₂, elevated O₃, and elevated O₃ + CO₂) and control plots arranged in a randomized complete block design with three replicates¹¹. FACE technology combines a trace gas monitoring system with a delivery system composed of blowers and vertical pipes placed around the perimeter of the forest stand to elevate ambient concentrations of O₃ and CO₂ (ref. 11). Ambient levels of CO₂ and O₃ are approximately 347 μl l⁻¹ and 37 nl l⁻¹, respectively. Elevated CO₂ and elevated O₃ treatments averaged 560 μl l⁻¹ and 52 nl l⁻¹, respectively, since the fumigation began. The experiment was initiated in the summer of 1997 with the planting of greenhouse-propagated tree seedlings at 1 × 1 m spacing, in the following combinations within three sections of each 30-m ring: (1) aspen (*Populus tremuloides* Michx.); (2) aspen and birch (*Betula papyrifera* Marsh.); and (3) aspen and maple (*Acer saccharum* Marsh.). Fumigation began in May 1998. In this investigation we focus only on soil carbon in the aspen and aspen–birch stands. Soils were sampled in 1997 before the start of fumigation with elevated CO₂ and O₃, and again in October 2001 when trees had grown to approximately 5 m height, forming a closed-canopy system in all but the elevated O₃ treatment where trees were smaller with less leaf production. Soils at the site are mixed, frigid, coarse loamy Alfic Haplorthods. The texture of the soil is sandy loam, underlain by a clay plough-layer, and then grading back into a sandy loam.

Soil carbon fractionation

Soil cores (10 cm diameter × 20 cm deep) were collected at 10 random locations within each community type in each plot, composited by community type within each plot, sieved to remove roots and other coarse materials, and oven-dried before analysis for total soil carbon and δ¹³C. Subsamples (40 g) of each soil were extracted with 150 ml of deionized water overnight to remove water-soluble carbohydrates. We then removed any remaining fine organic matter on a portion of the residual soil by flotation with 1.0 M NaCl, rinsing with deionized water, and then manually removing any remaining particulate organic matter (POM) under a compound microscope at 20 × magnification. We then performed acid hydrolysis with 6 M HCl on a 2-g subsample of POM-free soil under reflux for 8 h to isolate the most decay-resistant carbon for %C and δ¹³C. Analyses of %C and δ¹³C were performed on a Costech Elemental Combustion System 4010 connected to a ThermoFinnigan ConFloIII Interface and DeltaPlus Continuous Flow-Stable Isotope Ratio Mass Spectrometer (IRMS).

Soil incubation and respiration measurements

A separate subsample of sieved root-free soil was used to determine potential carbon availability by aerobic incubation. We placed 100 g of field moist soil into microlysimeters constructed from funnels, brought soils to approximately 60% field capacity, then placed soils into 1-litre incubation jars which were placed in an incubator at 25.0 ± 0.5 °C. At intervals that increased in length over time, we measured CO₂ accumulation in the headspace by withdrawing 50 μl from sealed jars fitted with septa and measured the

concentration using a Varian CP-3800 gas chromatograph with a thermal conductivity detector. Rates of CO₂ production were calculated by dividing concentrations by the time since jars had been sealed. On the same date, we collected samples of gas from the headspace of incubation jars for determining the $\delta^{13}\text{C}$ value of the accumulated CO₂. We injected 7 ml of headspace gas into He-flushed gas-tight LabCo exetainers, which were then placed into an autosampler and analysed on a FIN-MAT Gas Bench II connected to the IRMS. Soil respiration $\delta^{13}\text{C}$ values for control and elevated O₃ treatments were $-25.0 \pm 0.2\text{‰}$ across the incubation. For soils from elevated CO₂, soil respiration values initially were $-30.9 \pm 0.3\text{‰}$ and increased to $-28.8 \pm 0.2\text{‰}$ by day 281 of incubation. For the elevated O₃ + CO₂ treatment, values initially were $-31.1 \pm 0.4\text{‰}$ and -28.2 ± 0.2 at day 281 of incubation.

Stable isotope tracer, calculations and statistical analysis

To investigate the contribution of new carbon inputs to soil and microbial respiration in the elevated CO₂ and elevated O₃ + CO₂ plots, we used the $\delta^{13}\text{C}$ signature derived from the fossil-fuel fumigation gas mixed with ambient air ($-18.7 \pm 1.0\text{‰}$ compared to $-8.6 \pm 0.1\text{‰}$ for control and elevated O₃ plots). Subsequent fractionation by the plants produced leaf and root tissue with $\delta^{13}\text{C}$ values of approximately $-41.6 \pm 0.4\text{‰}$ in elevated CO₂ and elevated O₃ + CO₂ plots. Composite soil samples collected from all 12 plots before the experiment began in 1997 and from the three control plots in 2001 had similar ($P = 0.47$) $\delta^{13}\text{C}$ values, averaging $-26.7 \pm 0.2\text{‰}$. The proportional contribution of carbon derived from the fumigation gas (f) was calculated using the equation $f = (\delta_i - \delta_o)/(\delta_i - \delta_o)$, where δ_i refers to the isotopic composition ($\delta^{13}\text{C}$) of the soil organic carbon or soil respiration from the fumigated (elevated CO₂ or elevated O₃ + CO₂) plots, δ_o is the $\delta^{13}\text{C}$ of soil organic carbon or soil respiration from control plots, and δ_i is the $\delta^{13}\text{C}$ signature of the plant leaves and roots collected in 2001 from the fumigated plots, weighted equally and averaged across species as there were no significant differences found among the aspen and aspen–birch plots. We do not expect that the isotopic signature of the plants varied appreciably over the life of this experiment because fumigation was initiated when the trees were still seedlings. Because there were differences among treatments in the amount of carbon entering soils derived from fumigation (Fig. 1), we calculated the relative mass loss as the cumulative amount of carbon respired from each of these pools divided by the amount of soil carbon in that pool. Data were analysed using analysis of variance with general linear models for a split-plot randomized complete design using SAS 8.02 (Cary, North Carolina).

Received 25 February; accepted 12 September 2003; doi:10.1038/nature02047.

1. IPCC Climate Change 2001: Technical Summary (Report of the Intergovernmental Panel on Climate Change, IPCC Secretariat, Geneva, 2001).
2. Gregg, J. W., Jones, C. G. & Dawson, T. E. Urbanization effects on tree growth in the vicinity of New York City. *Nature* **424**, 183–187 (2003).
3. McLaughlin, S. B. & Downing, D. J. Interactive effects of ambient ozone and climate measured on growth of mature forest trees. *Nature* **374**, 252–254 (1995).
4. Chameides, W. L., Kasibhatla, P. S., Yienger, J. & Levy, H. I. Growth of continental-scale metro-agroplexes, regional ozone pollution, and world food production. *Science* **264**, 74–77 (1994).
5. Percy, K. E. *et al.* Altered performance of forest pests under CO₂ and O₃-enriched atmospheres. *Nature* **420**, 403–407 (2002).
6. Latest Findings on National Air Quality: 2002 Status and Trends (US Environmental Protection Agency).
7. Findlay, S., Carreiro, M., Krichik, V. & Jones, C. G. Effects of damage to living plants on leaf litter quality. *Ecol. Appl.* **6**, 269–275 (1996).
8. Coleman, M. D., Dickson, R. E., Isebrands, J. G. & Karnosky, D. F. Carbon allocation and partitioning in aspen clones varying in sensitivity to tropospheric ozone. *Tree Physiol.* **15**, 593–604 (1995).
9. Andersen, C. P. Source-sink balance and carbon allocation below ground in plants exposed to ozone. *New Phytol.* **157**, 213–228 (2003).
10. King, J. S. *et al.* Fine-root biomass and fluxes of soil carbon in young stands of paper birch and trembling aspen as affected by elevated atmospheric CO₂ and tropospheric O₃. *Oecologia* **128**, 237–250 (2001).
11. Dickson, R. E. *et al.* Forest Atmosphere Carbon Transfer and Storage (FACTS-II)—The Aspen Free-air CO₂ and O₃ Enrichment (FACE) Project: An Overview (Technical Report NC-214, USDA, Washington DC, 2000).
12. Leavitt, S. W., Follett, R. F. & Paul, E. A. Estimation of slow- and fast-cycling soil organic carbon pools from 6N HCl hydrolysis. *Radiocarbon* **38**, 231–239 (1996).
13. Paul, E. A. *et al.* Radiocarbon dating for determination of soil organic matter pool sizes and dynamics. *Soil Sci. Soc. Am. J.* **61**, 1058–1067 (1997).
14. Phillips, R., Zak, D. R., Holmes, W. E. & White, D. C. Microbial community composition and function beneath temperate trees exposed to elevated atmospheric carbon dioxide and ozone. *Oecologia* **131**, 236–244 (2002).
15. Larson, J., Zak, D. R. & Sinsabaugh, R. L. Extracellular enzyme activity beneath temperate trees growing under elevated carbon dioxide and ozone. *Soil Sci. Soc. Am. J.* **66**, 1848–1856 (2002).

Acknowledgements This research was supported by the US Department of Energy's Office of Science (BER: Program for Ecosystem Research and National Institute for Global Environmental Change), the USDA Forest Service (Northern Global Change and North Central Research Station), the National Science Foundation (DEB, DBI/MRI), and the USDA Natural Research Initiatives Competitive Grants Program. G. Hendry, K. Lewin, J. Nagey, D. Karnosky and J. Sober have been instrumental in the successful implementation of this long-term field experiment.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.M.L. (wmloya@mtu.edu).

Flying and swimming animals cruise at a Strouhal number tuned for high power efficiency

Graham K. Taylor, Robert L. Nudds* & Adrian L. R. Thomas

Zoology Department, University of Oxford, Tinbergen Building, South Parks Road, Oxford OX1 3PS, UK

* Present address: School of Biology, University of Leeds, L. C. Miall Building, Clarendon Way, Leeds LS2 9JT, UK

Dimensionless numbers are important in biomechanics because their constancy can imply dynamic similarity between systems, despite possible differences in medium or scale¹. A dimensionless parameter that describes the tail or wing kinematics of swimming and flying animals is the Strouhal number¹, $St = fA/U$, which divides stroke frequency (f) and amplitude (A) by forward speed (U)^{2–8}. It is known to govern a well-defined series of vortex growth and shedding regimes for airfoils undergoing pitching and heaving motions^{6,8}. Propulsive efficiency is high over a narrow range of St and usually peaks within the interval $0.2 < St < 0.4$ (refs 3–8). Because natural selection is likely to tune animals for high propulsive efficiency, we expect it to constrain the range of St that animals use. This seems to be true for dolphins^{2–5}, sharks^{3–5} and bony fish^{3–5}, which swim at $0.2 < St < 0.4$. Here we show that birds, bats and insects also converge on the same narrow range of St , but only when cruising. Tuning cruise kinematics to optimize St therefore seems to be a general principle of oscillatory lift-based propulsion.

Experiments with isolated pitching or heaving foils have measured extremely high peak propulsive efficiencies within the interval $0.2 < St < 0.4$ (modal peak at $St \approx 0.3$)^{3–7}. In this range, the propulsive efficiency (defined as the ratio of aerodynamic power output to mechanical power input) can be as high as 70% (ref. 7) or even 80% (ref. 6). Optimal St depends subtly on kinematic parameters including geometric angle of attack, amplitude-to-chord ratio, airfoil section and phase of motion^{6–8} but, for any given motion, efficiency is usually high ($>60\%$) over a range narrower than $0.2 < St < 0.4$ (refs 3–7). For example, measured efficiency can plummet from 80% at $St = 0.27$ to 10% at $St = 0.09$ (ref. 6) and also drops off at higher St , albeit more gently^{7,8}. Measured propulsive efficiency usually peaks when the kinematics result in maximum amplification of the shed vortices in the wake and an average velocity profile equivalent to a jet^{3,4,6}.

Theoretical treatments of flapping wings^{3,4,6,8} further confirm the empirical result that St tightly constrains propulsive efficiency. In fact, St is bound to affect aerodynamic force coefficients and propulsive efficiency, because it defines the maximum aerodynamic angle of attack and the timescales associated with the growth and shedding of vortices, which are the source of aerodynamic force production^{8,9}. Natural selection is expected to favour wing kinematics that combine high propulsive efficiency with a high aerodynamic force coefficient. These need not peak at identical St , but can do for certain motions^{7,8}; if not, selection should optimize the trade-off. Propulsive efficiency may be the more important selection pressure in cruising, whereas high aerodynamic force coefficients may be more important in accelerations, slow locomotion or hovering. In cruising flight or swimming, we therefore predict that St will be tuned for high propulsive efficiency.

This suggestion has already been made for cruising fish and dolphins^{2–5}, which operate within the range $0.2 < St < 0.4$, and the principle is considered so general for swimming animals that it has even been used to predict the speeds of extinct ichthyosaurs¹⁰. Whereas the fluid dynamic results described above^{3–8} refer to

isolated flapping foils, swimming animals usually beat their tail in the wake of their own body. Non-streamlined bodies shed vortices at a natural frequency (also defined by a Strouhal number), which the tail can tune into to regain energy lost into the wake^{5,11–13}. As this will complicate selection of St , swimming animals may be less well modelled by isolated flapping foils than are flying animals (whose wings do not operate in their body's wake, thereby avoiding competition between a natural and a forced shedding mode). Similar complications may arise where favourable interactions occur between the wings and wake following stroke reversal, or where the fore and hind wings interact. This will be especially problematic in slow-flying insects, and may be important for insects generally, but the paucity of kinematic data for cruising insects makes this difficult to test.

The results for isolated flapping foils consider only pitching and heaving motions, such as those used by swimming animals^{3–8}. The root-flapping motions of flying animals seem not to have been tested. We therefore subjected a hinged flat plate to heaving or root-flapping motions to check whether these motions induced the same vortex growth and shedding regimes at similar St . We equate stroke amplitude with wingtip excursion, which makes sense because the wake is bounded by vorticity shed at the wingtips and the aerodynamics are dominated by the faster-moving outboard wing elements. This simple definition is also directly equivalent in

heaving, pitching and root-flapping motions. We found that heaving generated larger vortices than root-flapping (presumably because average velocity was higher across the span), but otherwise the same four distinct wake regimes occurred at similar St (Fig. 1). By analogy with swimming animals, we therefore predicted that flying animals would have been selected to operate in the same range $0.2 < St < 0.4$ for efficient cruising.

To test this prediction, we identified 16 studies that reported wingbeat frequency, stroke amplitude and flight speed^{14–29} in steady cruising flight. All of the analysis was done with Matlab 6.5. If wingtip excursion A was not directly measured, we approximated A as $A \approx b \sin \theta / 2$, where b is span and θ is dorsoventral stroke angle. This allowed us to calculate St for 42 species (Fig. 2) from 14 bird ($n = 22$), 6 bat ($n = 18$) and 2 insect ($n = 2$) families. Insect flight research has tended to focus on slow or hovering flight, and further data on cruising insects are clearly required before generalizing to other insects, especially as cruising power efficiency may not be a strong selection pressure in many species. In most cases, we used mean or modal values collected from animals flying naturally in the wild. In wind tunnel studies in which the animal was forced to fly at pre-determined speeds, St was calculated for the reported preferred flying speed^{22,23,26,27}. We excluded studies in which the animal was otherwise confined and liable to be flying slowly.

Figure 2 clearly shows that swimming and flying animals all

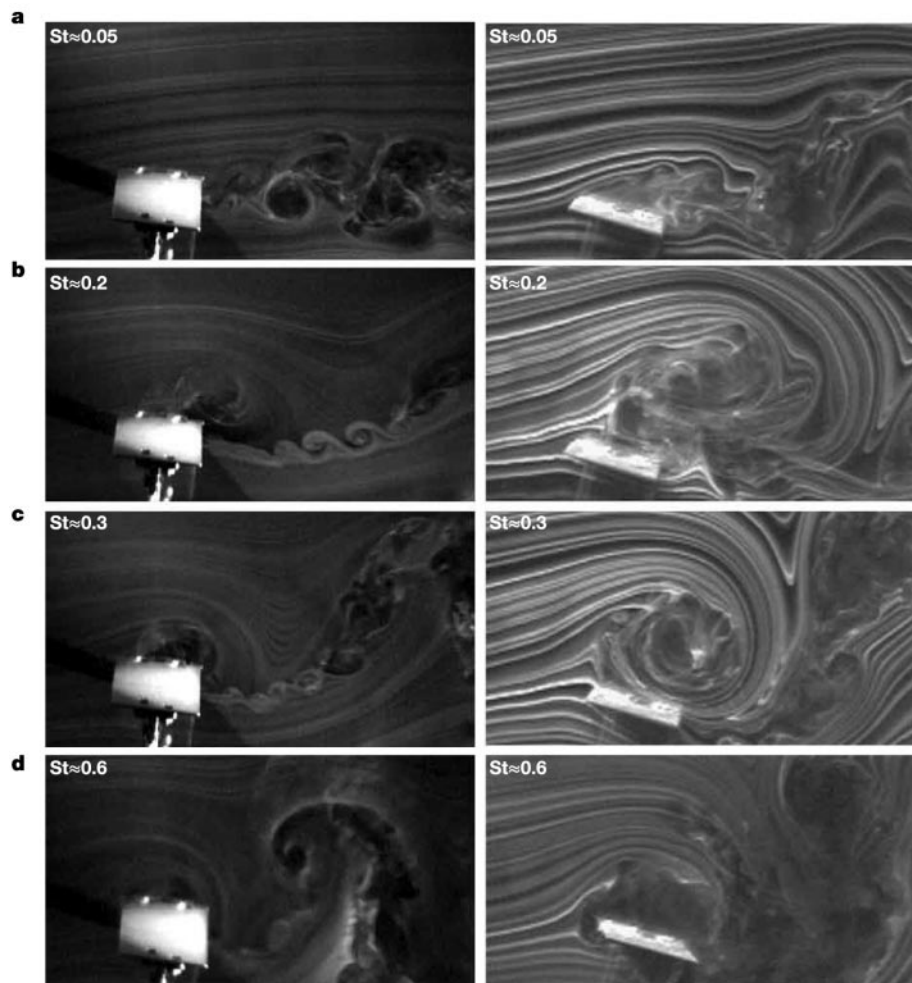


Figure 1 Wake structures for root-flapping and heaving hinged flat plates at varying St . Left panels, root-flapping motion; right panels, heaving motion. Amplitude, twice wing chord; static angle of attack, 15° ; flow speed, 1.5 ms^{-1} ; smoke wire visualizations made at end of downstroke. **a**, For $St < 0.10$, flow separates at the sharp leading-edge, but no discrete vortex forms. **b**, For $0.10 < St < 0.25$, a leading-edge vortex forms but is shed

before the downstroke ends. **c**, For $0.25 < St < 0.45$, the leading-edge vortex is shed as the downstroke ends. **d**, For $St > 0.45$, trailing edge separation produces a characteristic mushroom-shaped wake. At higher St , the wing collides with shed vorticity on the upstroke, giving an energetically inefficient mode.

operate at similar St . Does this reflect selection to constrain St , or is it merely coincidental? To answer this, we first provide two independent confirmations that St is tightly constrained in cruising flight. The first confirmation comes from considering how St varies when an animal is forced to fly other than at its preferred speed. St varies more in four individual zebra finches (*Taenopygia guttata*)²⁷ forced to fly between 4 and 14 ms^{-1} than across all 42 species flying at their preferred speeds (Fig. 2). Even excluding the lowest forced flight speeds (which are lower than any preferred flight speed), the standard deviation (s.d.) of St for the zebra finch across speeds is more than twice that for all 42 species flying at their preferred speeds. Flight speed affects St so strongly because wingbeat frequency and amplitude are tightly constrained, presumably by physiology and morphology. This makes it even more notable that, when only cruise performance is considered, a sample spanning five orders of body mass and three independent evolutionary origins of flight shows less than half the variation in St found in four individuals of a single species forced to fly at different speeds.

A Monte Carlo analysis provides a second confirmation that St is tightly constrained in cruising flight (Fig. 3). As frequency, speed and amplitude all scale with body mass (m), if wing kinematics are indeed tuned to optimize St , their residual variation should co-vary appropriately to constrain St . We therefore regressed $\log(f)$, $\log(U)$ and $\log(A)$ separately against $\log(m)$ and calculated the fitted values for each species. We then randomly allocated 1 of the 42 residuals from each of the regressions to each species without replacement and calculated the s.d. of St in the resulting sample ($n = 42$). We repeated this procedure 50,000 times.

Figure 3 shows the distribution of the s.d. of resampled St . Regressions and residual plots (Supplementary Fig. S1) showed sufficient homoscedasticity for the procedure to be valid. Only 53 of the 50,000 randomized combinations of residuals ($<0.11\%$) had a lower s.d. of St than the original sample. The actual s.d. of St (0.10) was therefore significantly smaller than expected by chance

($P = 0.001$) from the allometry of f , U and A . This is very strong evidence that St is tightly constrained during cruising flight. It would not be unexpected to find St tightly constrained in similar species, because similar morphological and physiological constraints usually produce dynamic similarity¹; however, it is surprising to find St tightly constrained in this morphologically and physiologically disparate sample, unless St is constrained by one unifying factor— aerodynamics.

The propulsive efficiency of an isolated flapping foil usually peaks at $St \approx 0.3$ (refs 3–7), therefore we used a two-tailed t -test to check that actual mean $St = 0.29$ did not differ significantly ($P = 0.136$) from this (after square-root transforming the data to remove skew arising because St is a ratio). The data are not normally distributed even after transformation (Supplementary Fig. S2), but the t -test is robust to large deviations from normality and, as the 95% confidence interval ($0.25 < St < 0.31$) fell comfortably inside the optimal range $0.2 < St < 0.4$, we can be reasonably sure that the lack of any statistically significant difference is not merely an artefact of low statistical power. Median $St = 0.25$ (inter-quartile range 0.20–0.35) was less than mean St because of the positive skew, but still did not differ significantly from the expected optimum (sign test, $P = 0.081$). St therefore seems to have converged on the expected optimum $St \approx 0.3$ in cruising flight, with about 75% of species falling in the range $0.19 < St < 0.41$. In addition, the fit is remarkably tight given the much larger variation in St found in a single species forced to fly at different speeds.

Because the assumptions of parametric analysis of variance are unlikely to hold for our data set, we used a Kruskal–Wallis test to compare St among taxonomic classes. The test just failed to attain statistical significance ($P = 0.054$), mainly because the $n = 2$ sample size for insects limited its efficiency. Using a Wilcoxon rank sum test to compare directly birds and bats showed instead that St is significantly lower in birds than in bats (two-tailed, $P = 0.020$). We then dropped all families of $n = 1$ and used a

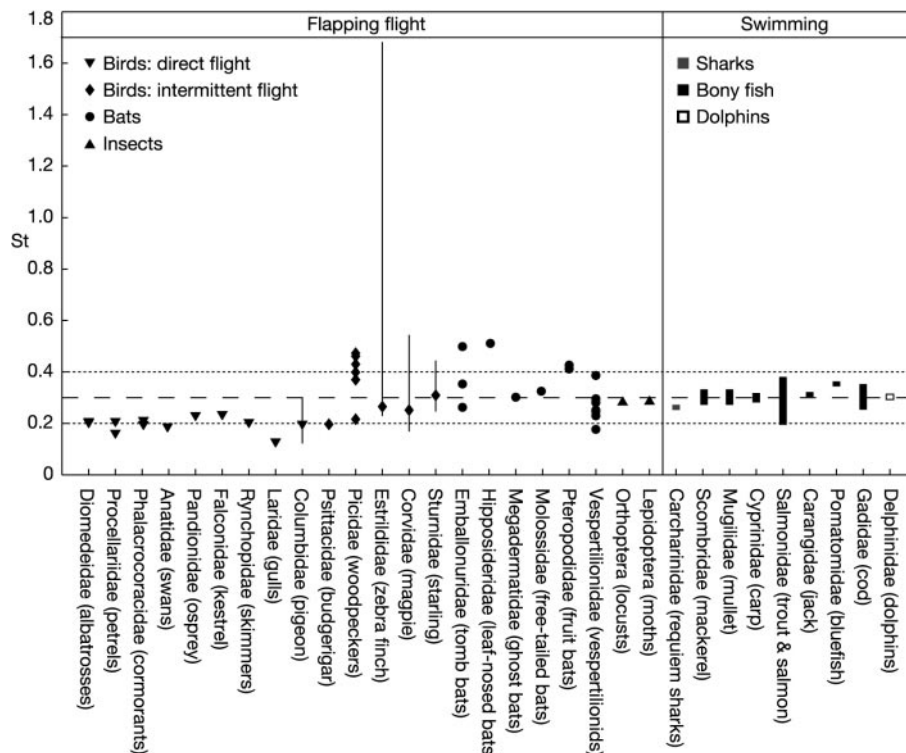


Figure 2 Strouhal number for 42 species of birds, bats and insects in unconfined, cruising flight. Published ranges^{3,4} of St in cruising fish and dolphins are included for comparison. Dotted lines mark the range $0.2 < St < 0.4$, in which propulsive efficiency usually

peaks; dashed line marks the modal peak at $St = 0.3$. Unbroken lines indicate the range of variation in St across other non-zero flight speeds, where such data exist.

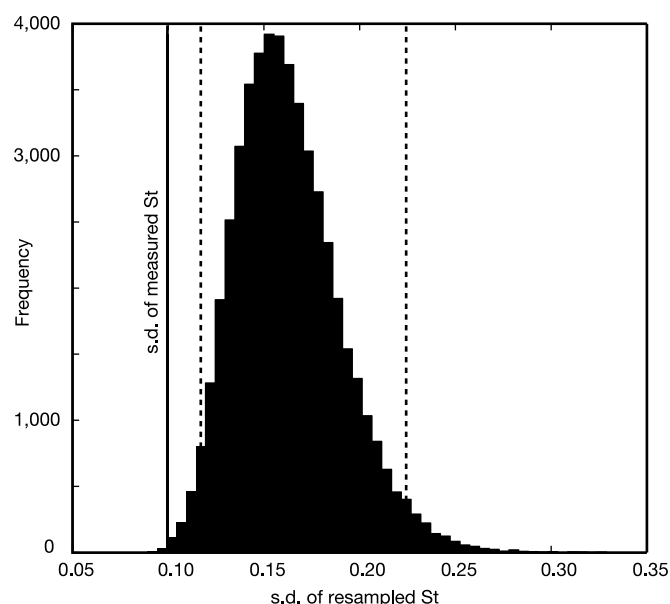


Figure 3 Histogram from Monte Carlo analysis recalculating St for 50,000 iterations randomizing the order of residuals from regressing $\log(f)$, $\log(U)$ and $\log(A)$ against $\log(m)$. Dotted lines denote the 95% confidence interval for the s.d. of the sample expected by chance from the regressions; the actual s.d. was significantly less than expected by chance ($P = 0.001$).

Kruskal–Wallis test to look for interfamilial variation within birds and bats. St varied significantly between the four remaining bird (two-tailed $P = 0.036$) and the three remaining bat (two-tailed, $P = 0.032$) families. *A posteriori* comparisons between the mean ranks using Tukey's honestly significant difference procedure ($\alpha = 0.05$) were too conservative to identify which differences were significant (the procedure forfeits sensitivity for individual differences to control the overall risk of type I error). We therefore used Tukey's least significant difference procedure ($\alpha = 0.05$), which is more liberal but indicated that St was significantly higher for woodpeckers than for petrels, and significantly lower for vesperilionid bats than for fruit bats.

Birds using intermittent flight (median $St = 0.34$, inter-quartile range 0.25–0.43, $n = 10$) have significantly higher St (Wilcoxon rank sum test, $P < 0.001$) than birds using direct flight (median $St = 0.20$, inter-quartile range 0.19–0.21, $n = 12$). Direct fliers occupy the lower extreme of the range $0.2 < St < 0.4$ (Fig. 2), where leading-edge separation is weak or non-existent for classical airfoils⁶. Intermittent fliers seem to occupy the upper end of this range, where leading-edge separation is more likely to occur (because maximum angle of attack increases with St), but as the average flight speeds used to calculate St underestimate instantaneous speed during flapping, their true instantaneous St will be lower. Unlike bats and insects, birds have wings that are well shaped for maintaining attached flows, which offer the highest possible lift-to-drag ratios by minimizing the kinetic energy losses associated with separated flows.

Intermittent flight could further help to maintain attached flows by lowering instantaneous St , but when might this be important? Allometry implies that U should scale as $m^{1/6}$ and A as $m^{1/3}$, thus St will be scale-invariant if f scales as $m^{-1/6}$. It has been found³⁰ that f actually scales as $m^{-3/11}$, implying that St should scale as $m^{-7/66}$ ($m^{-0.11}$). Regressing $\log(St)$ against $\log(m)$ for the birds in our data set (Supplementary Fig. S3) confirms that St scales as roughly $m^{-0.12}$ ($n = 22$, $P = 0.003$). Constraints on wingbeat frequency may therefore mean that smaller birds have intrinsically higher St , which could make intermittent flight important for lowering St to maintain attached flows when cruising. As leading-edge separation

is probably inevitable at the very high St associated with slow flight (Fig. 2), we predict that separated flows will be normal for birds in very slow flight.

We have provided several independent confirmations that flying animals both operate within a narrow range of St when cruising and have converged on the optimum range $0.2 < St < 0.4$ expected for high propulsive efficiency and also used by cruising fish and dolphins. Cruising flight and swimming speeds therefore may be predicted as $U = fA/St$, where $St \approx 0.3$. As a simple rule of thumb, a cruising animal will move at a speed just over three times the product of its stroke frequency and amplitude. Conversely, a flapping micro-air vehicle of 15-cm span cruising with a 90° stroke angle (10-cm stroke amplitude) at 10 ms^{-1} should attain peak propulsive efficiency at a wingbeat frequency of just over 30 Hz. Although the precise optimum St will vary with wing morphology and kinematics, this scaling seems to be general. It applies to animals moving through air and through water. It broadly applies whether the propulsion is driven by the wings or tail. It applies (in our data set) to animals ranging in size from moths to dolphins. If there are swimming or flying organisms on other planets, then we predict that it should apply to them too. □

Received 17 July; accepted 19 August 2003; doi:10.1038/nature02000.

- Alexander, R. M. *Principles of Animal Locomotion* (Princeton Univ. Press, Princeton, 2003).
- Rohr, J. J. et al. *Observations of Dolphin Swimming Speed and Strouhal Number*. Space and Naval Warfare Systems Center Technical Report No. 1769 (Space and Naval Warfare Systems Center, San Diego, 1998).
- Triantafyllou, M. S., Triantafyllou, G. S. & Gopalkrishnan, R. Wake mechanics for thrust generation in oscillating foils. *Phys. Fluids A* **3**, 2835–2837 (1991).
- Triantafyllou, G. S., Triantafyllou, M. S. & Grosenbaugh, M. A. Optimal thrust development in oscillating foils with application to fish propulsion. *J. Fluids Struct.* **7**, 205–224 (1993).
- Triantafyllou, M. S., Triantafyllou, G. S. & Yue, D. K. P. Hydrodynamics of fishlike swimming. *Annu. Rev. Fluid Mech.* **32**, 33–53 (2000).
- Anderson, J. M., Streitlien, K., Barrett, D. S. & Triantafyllou, M. S. Oscillating foils of high propulsive efficiency. *J. Fluid Mech.* **360**, 41–72 (1998).
- Read, D. A., Hover, F. S. & Triantafyllou, M. S. Forces on oscillating foils for propulsion and maneuvering. *J. Fluids Struct.* **17**, 163–183 (2003).
- Wang, Z. J. Vortex shedding and frequency selection in flapping flight. *J. Fluid Mech.* **410**, 323–341 (2000).
- Huang, R. F., Wu, J. Y., Jeng, J. H. & Chen, R. C. Surface flow and vortex shedding of an impulsively started wing. *J. Fluid Mech.* **441**, 265–292 (2001).
- Motani, R. Scaling effects in caudal fin propulsion and the speed of ichthyosaurs. *Nature* **415**, 309–312 (2002).
- Bandyopadhyay, P. R., Castano, J. M., Nedderman, W. H. & Donnelly, M. J. Experimental simulations of fish-inspired unsteady vortex-dynamics on a rigid cylinder. *J. Fluids Eng.* **122**, 219–238 (2000).
- Wolfgang, M. J., Anderson, J. M., Grosenbaugh, M. A., Yue, D. K. P. & Triantafyllou, M. S. Near-body flow dynamics in swimming fish. *J. Exp. Biol.* **202**, 2303–2327 (1999).
- Gopalkrishnan, R., Triantafyllou, M. S., Triantafyllou, G. S. & Barrett, D. Active vorticity control in a shear flow using a flapping foil. *J. Fluid Mech.* **274**, 1–21 (1994).
- Withers, P. C. & Timko, P. L. The significance of ground effect to the aerodynamic cost of flight and energetics of the black skimmer (*Rhynchops nigra*). *J. Exp. Biol.* **70**, 13–26 (1977).
- Baker, P. S. & Cooter, R. J. The natural flight of the migratory locust, *Locusta migratoria* L. I. Wing movements. *J. Comp. Physiol. A* **131**, 79–87 (1979).
- Baker, P. S., Gewecke, M. & Cooter, R. J. The natural flight of the migratory locust, *Locusta migratoria* L. III. Wing-beat frequency, flight speed and attitude. *J. Comp. Physiol.* **141**, 233–237 (1981).
- Scholey, K. D. *Developments in Vertebrate Flight: Climbing and Gliding of Mammals and Reptiles and the Flapping Flight of Birds*. Thesis, Univ. Bristol (1983).
- Videler, J. J., Groenewegen, A., Gnodde, M. & Vossebelt, G. Indoor flight experiments with trained kestrels II. The effect of added weight on flapping flight kinematics. *J. Exp. Biol.* **134**, 185–199 (1988).
- Pennycuik, C. J. Span-ratio analysis used to estimate effective lift:drag ratio in the double-crested cormorant *Phalacrocorax auritus* from field observations. *J. Exp. Biol.* **142**, 1–15 (1989).
- Dudley, R. & DeVries, P. J. Flight physiology of migrating *Urania fulgens* (Uranidae) moths: kinematics and aerodynamics of natural free flight. *J. Comp. Physiol. A* **167**, 145–154 (1990).
- Pennycuik, C. J. Predicting wingbeat frequency and wavelength of birds. *J. Exp. Biol.* **150**, 171–185 (1990).
- Tobalske, B. W. & Dial, K. P. Neuromuscular control and kinematics of intermittent flight in budgerigars (*Melopsittacus undulatus*). *J. Exp. Biol.* **187**, 1–18 (1994).
- Tobalske, B. W. Neuromuscular control and kinematics of intermittent flight in the European starling (*Sturnus vulgaris*). *J. Exp. Biol.* **198**, 1259–1273 (1995).
- Pennycuik, C. J. Wingbeat frequency of birds in steady cruising flight: new data and improved predictions. *J. Exp. Biol.* **199**, 1613–1618 (1996).
- Tobalske, B. W. Scaling of muscle composition, wing morphology, and intermittent flight behavior in woodpeckers. *Auk* **113**, 151–177 (1996).
- Tobalske, B. W. & Dial, K. P. Flight kinematics of black-billed magpies and pigeons over a wide range of speeds. *J. Exp. Biol.* **199**, 263–280 (1996).
- Tobalske, B. W., Peacock, W. L. & Dial, K. P. Kinematics of flap-bounding flight in the zebra finch over a wide range of speeds. *J. Exp. Biol.* **202**, 1725–1739 (1999).

28. Pennycuik, C. J. Speeds and wingbeat frequencies of migrating birds compared with calculated benchmarks. *J. Exp. Biol.* **204**, 3283–3294 (2001).
29. Bullen, R. D. & McKenzie, N. L. Scaling bat wingbeat frequency and amplitude. *J. Exp. Biol.* **205**, 2615–2626 (2002).
30. Rayner, J. M. V. Form and function in avian flight. *Curr. Ornithol.* **5**, 1–45 (1987).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Johnson for comments on the statistical analysis. This work was funded by grants from the Biotechnology and Biological Sciences Research Council and a Christopher Welch Scholarship. G.K.T. is a Royal Commission for the Exhibition of 1851 Research Fellow and Weir Junior Research Fellow at University College, Oxford.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to G.K.T. (graham.taylor@zoo.ox.ac.uk).

New frog family from India reveals an ancient biogeographical link with the Seychelles

S. D. Biju^{1,2*} & Franky Bossuyt^{2*}

¹Tropical Botanic Garden and Research Institute, Palode, Thiruvananthapuram, 695562 Kerala, India

²Biology Department, Unit of Ecology & Systematics, Vrije Universiteit Brussel, Pleinlaan 2, B-1050 Brussels, Belgium

* The authors contributed equally to this work

About 96% of the more than 4,800 living anuran species¹ belong to the Neobatrachia or advanced frogs^{2–4}. Because of the extremely poor representation of these animals in the Mesozoic fossil record, hypotheses on their early evolution have to rely largely on extant taxa^{5–7}. Here we report the discovery of a burrowing frog from India that is noticeably distinct from known taxa in all anuran families. Phylogenetic analyses of 2.8 kilobases of mitochondrial and nuclear DNA unambiguously designate this frog as the sister taxon of Sooglossidae, a family exclusively occurring on two granitic islands of the Seychelles archipelago⁸. Furthermore, molecular clock analyses⁹ uncover the branch leading to both taxa as an ancient split in the crown-group Neobatrachia. Our discovery discloses a lineage that may have been more diverse on Indo-Madagascar in the Cretaceous period, but now only comprises four species on the Seychelles and a sole survivor in India. Because of its very distinct morphology and an inferred origin that is earlier than several neobatrachian families¹⁰, we recognize this frog as a new family.

Amphibia L., 1758
Lissamphibia Haeckel, 1866
Anura Rafinesque, 1815
Neobatrachia Reig, 1958
Nasikabatrachidae fam. nov.
Nasikabatrachus gen. nov.

Nasikabatrachus sahyadrensis gen. et sp. nov.

Etymology. Nasika (Sanskrit) meaning nose, batrachus meaning frog, and Sahyadri, being synonymous for the Western Ghats (the hills along the west coast of the Indian subcontinent).

Holotype. Bombay Natural History Society (BNHS; Mumbai), BNHS 4202, an adult female, snout–vent length 70.1 mm, collected July 2000 by S.D.B. (Fig. 1a).

Type locality. Disturbed secondary forest near a cardamom planta-

tion at Kattappana (09° 45' N, 77° 05' E, altitude approximately 900 m), Idukki district, Kerala, Western Ghats, India.

Diagnosis. The diagnosis is valid for the family, genus and species. A relatively large frog with a bloated general appearance, smooth skin and an overall black coloration dorsally and dark grey ventrally; the head (Fig. 1b) is pointed and short relative to the body; the snout has a distinct white protrusion. The eyes are small with a rounded, horizontal pupil; no apparent tympanum; the forelimbs are short, the hands (Fig. 1c) are rudimentarily webbed, the tips of fingers are rounded, without disks. The hindlimbs are short, feet (Fig. 1d) are about 3/4 webbed, and the tips of toes are rounded, without disks. A large, white inner metatarsal tubercle is present on both feet (detailed measurements of external morphology are provided as Supplementary Information).

The skeleton (Fig. 1e) is characteristic of a burrowing frog and displays bones with a well-calcified cortical area, a skull with strongly ossified neurocranial and dermal elements (Fig. 1f), a short tibiae and fibulae, strong and short tibiofibular bones, and

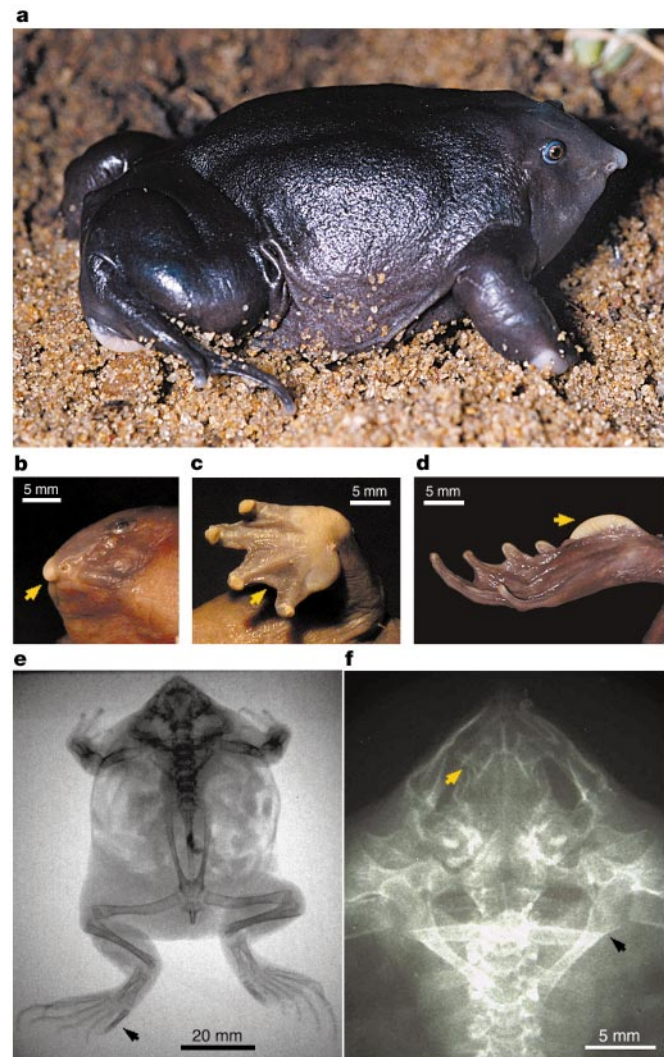


Figure 1 Holotype of *Nasikabatrachus sahyadrensis*. **a**, *Nasikabatrachus sahyadrensis* in life. **b**, Detail of head, showing slender mouth and distinct protrusion on snout. **c**, Detail of hand, showing rudimentary webbing. **d**, Detail of foot showing the large, white inner metatarsal tubercle. **e**, X-ray photograph showing strongly calcified bones. The arrow indicates the prehallux. **f**, X-ray photograph showing strongly ossified skull and pectoral girdle. The yellow arrow indicates the presumed neopalatine bone; the black arrow indicates the coracoid, the lateral end of which is wider than the medial.

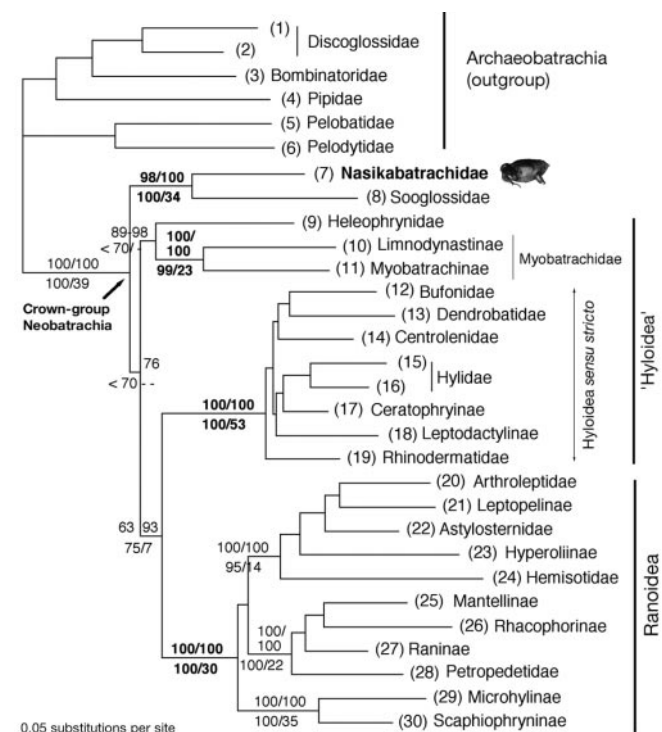


Figure 2 Bayesian consensus phylogram for the analysis of 2,325 bp of DNA. Numbers before (sub)familial names correspond to species as given in the Methods section. Numbers on branches indicate Bayesian posterior probabilities and metaGA posterior branch support values (above), respectively, and maximum parsimony bootstrap (10,000 replicates) values and decay indices (below), respectively. The sister group relationship of Nasikabatrachidae with Sooglossidae from the Seychelles is very strongly supported. Although the exact order of early divergences remains tentative, the strong support for several groups (clade support indicated in bold) implies that the Nasikabatrachidae/Sooglossidae lineage is one of the basal-most splits in Neobatrachia.

a well developed and highly calcified prehallux. The following osteological characters were tentatively interpreted from the X-rays: neobatrachian synapomorphies include fusion of the third carpal to other carpals, presence of a neopalatine bone (Fig. 1f) and absence of a parahyoid bone. The pectoral girdle with slender

Table 1 Bayesian divergence time estimates

Divergence	Time estimate (Myr ago)
Origin of Ranoidea and Hyloidea s.s.	152 (108, 202)
Origin of Heleophrynidae and Myobatrachidae	150 (109, 198)
Origin of Sooglossidae/Nasikabatrachidae lineage	178 (131, 233)
Sooglossidae–Nasikabatrachidae split	131 (93, 177)

See Methods and Supplementary Information for a description of the calibrations used for estimating node times. Numbers in parentheses represent 95% credibility intervals.

coracoids—the lateral end of which is wider than the medial (Fig. 1f)—excludes Nasikabatrachidae from Ranoidea³; a bony element situated at the medio-distal aspect of the tibial at the level of the third metatarsal, if homologous with the os sesamoides tarsale, links the new family to Sooglossidae³. The combination of the above external and skeletal characters makes Nasikabatrachidae distinct from all anuran families. In particular, the new lineage differs from its sister group Sooglossidae in many aspects, some of the most obvious being the absence of toe disks, a much larger size and the numerous adaptations to a burrowing lifestyle described above.

The phylogenetic position of Sooglossidae has been a point of debate for several decades, and they have been placed variously in Ranoidea¹¹, Hyloidea⁶ or in a trichotomy with the former⁵. Regardless of the optimality criteria used, our mitochondrial DNA, nuclear DNA and combined DNA analyses (see Supplementary Information for analyses of individual genes) strongly support Hyloidea *sensu stricto* and Ranoidea clades, and three additional main lineages: Myobatrachidae, Heleophrynidae and the Sooglossidae/Nasikabatrachidae clade (Fig. 2). In the absence of rate constancy, we used a relaxed molecular clock⁹ to estimate their divergence times under different possible topologies (derived from the Bayesian posterior probabilities), using multiple calibration points. All analyses resulted in similar dating estimates indicating that the major neobatrachian lineages originated relatively rapidly in the Middle/Late Jurassic and Early Cretaceous periods (Table 1). Around that time, the Gondwanan supercontinent broke up into two landmasses—western Gondwana (Africa and South America) and eastern Gondwana (Australia, Antarctica and Indo-Madagascar)—which rapidly disintegrated further into their respective components¹². These geological events probably isolated the stem group leading to the Nasikabatrachidae/Sooglossidae clade on the Indo-Madagascan fragment of eastern Gondwana (Fig. 3).

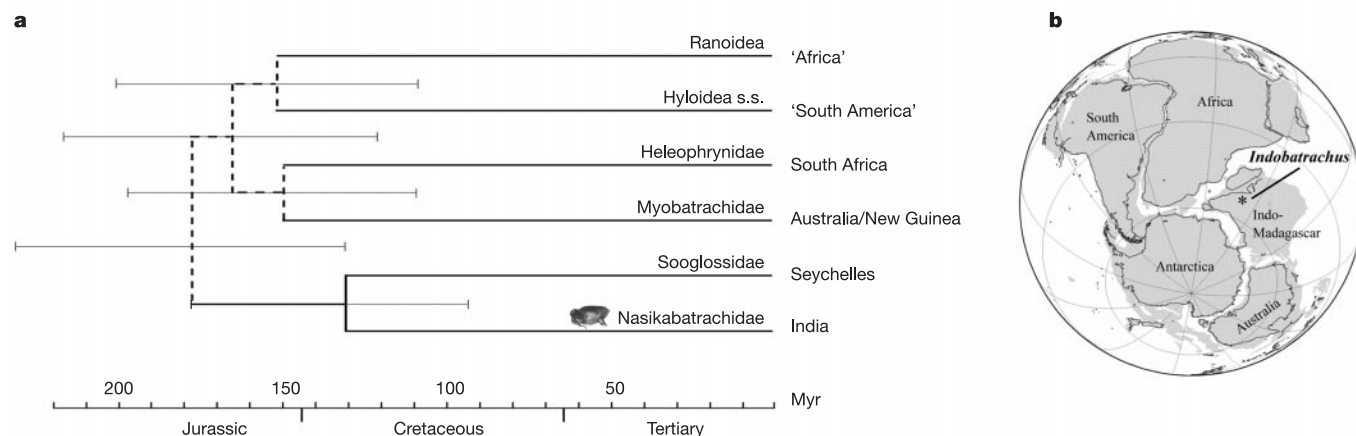


Figure 3 Early evolution of advanced frogs. **a**, Evolutionary tree showing estimates of divergence time for the major lineages, estimated from 1,443 bp of nuclear gene fragments. Analyses were performed on all taxa, but only basal stocks of modern Neobatrachia are shown. Dashed lines indicate uncertain phylogenetic relationships. Thin horizontal lines denote the 95% credibility interval for the corresponding node time. For

each lineage, the current distribution is indicated. Ranoidea and Hyloidea are now wide ranging, but their origin is associated with Africa and South America, respectively⁷. s.s., *sensu stricto*. **b**, Tectonic reconstruction of Gondwana 130.0 Myr ago. The locality of the fossil *Indobatrachus* from the Eocene epoch is indicated with an asterisk.

The endemism of Sooglossidae on the Seychelles has been a constant biogeographical mystery^{5,6,8}. One hypothesis postulates that this family is derived from an extant or extinct ancestral stock on Africa^{11,13}. Alternatively, it has been proposed that Sooglossidae were present on Indo-Madagascar⁶ during its trans-Tethys drift, and subsequently became extinct on India and Madagascar (if they had been present there). The phylogenetic position of Nasikabatrachidae, in combination with our dating estimates, clearly favour the second hypothesis. Furthermore, our analyses indicate that Nasikabatrachidae and Sooglossidae diverged well before the break-up of India and the Seychelles at the Cretaceous/Tertiary boundary (Table 1). Our findings thus uncover the existence of an early stock of advanced frogs, which consisted of at least two major lineages on drifting Indo-Madagascar. It is possible that this assemblage was composed of a much greater diversity during the Cretaceous period on parts of these Gondwanan fragments. Offshoots of this clade may even have existed on Australia and Antarctica, as some tectonic models propose a connection of the latter with Indo-Madagascar until the late Cretaceous period¹⁴.

It has already been suggested that India played a significant role in the passage of amphibians to southeast Asia^{4,15,16}. The phylogenetic pattern revealed here provides independent evidence for this land-mass being an important setting for the evolution of advanced frogs. There are no Mesozoic fossil records that testify to the existence of this early assemblage on eastern Gondwana. The only possible palaeontological evidence of a member of this group is *Indobatrachus pusillus* (Owen, 1847) from the Eocene epoch of India. This species has been regarded as a myobatrachid^{4,17,18} or sooglossid^{18,19}, but its osteology is still not completely known¹⁷. Comparison of additional fossil material with extant species of the Nasikabatrachidae/Sooglossidae clade may therefore provide new insights on the phylogenetic position of *Indobatrachus*.

Mesozoic fossils sometimes have important implications for understanding the early history of vertebrate groups^{20,21}. Yet, there are still very few palaeontological findings that have contributed to our knowledge on the origin of advanced frogs. Our discovery of an ancient extant frog lineage in India discloses a clade that significantly adds to the understanding of early neobatrachian biogeography. However, knowledge on how extensively the Indo-Madagascan lineage has radiated (both in the biogeographical and phylogenetic sense) will have to await the recovery of fossils from the Cretaceous period of Gondwanan landmasses. □

Methods

Osteological analyses

In order to perform a preliminary, non-destructive analysis of skeletal characters of the holotype, we made several X-ray photographs. We used a Philips Optimus M200 X-ray system and image intensifier (41 kV, 800 mA). Images were recorded digitally using a Redlake motion Pro high-resolution camera. Furthermore, an MPG 65 generator and a RSN 620 X-ray tube (General Electric) (56 kV, 200 mA) were used to study cranial elements in more detail.

Taxon selection and DNA methods

Our data set includes the following 30 species, sampled from all continental regions (including Madagascar and the granitic Seychelles) and forming a fair representation of the major anuran lineages recognized. Numbers correspond to (sub)familial names in Fig. 2: (1) *Alytes obstetricans boscai*, (2) *Discoglossus pictus*, (3) *Bombina orientalis*, (4) *Pipa pipa*, (5) *Pelobates cultripes*, (6) *Pelodytes punctatus*, (7) *Nasikabatrachus sahyadrensis*, (8) *Nesomantis thomasseti*, (9) *Heleophryne purcelli*, (10) *Limnodynastes salmini*, (11) *Myobatrachus gouldii*, (12) *Bufo melanostictus*, (13) *Dendrobates auratus*, (14) *Centrolene prosoblepon*, (15) *Hyla arenicolor*, (16) *Phrynohyas venulosa*, (17) *Ceratophrys ornata*, (18) *Leptodactylus melanonotus*, (19) *Rhinoderma darwini*, (20) *Arthroleptis variabilis*, (21) *Leptopelis kivuensis*, (22) *Trichobatrachus robustus*, (23) *Hyperolius* sp., (24) *Hemistis marmoratus*, (25) *Boophis xerophilus*, (26) *Philautus wynaadensis*, (27) *Meristogenys kinabaluensis*, (28) *Petropedetes cf. parkeri*, (29) *Microhyla ornata*, (30) *Scaphiophryne marmorata*. After preliminary analyses using salamanders as outgroups, Archaeobatrachia (species 1 to 6) were selected as outgroup taxa because of the evidence (osteological and molecular) for monophyly of Neobatrachia (including Nasikabatrachidae).

Two mitochondrial DNA gene fragments—covering part of 12S ribosomal RNA, complete transfer RNA^{Val} and part of 16S rRNA—were sequenced, which together yielded a data matrix of 1,329 base pairs (bp). Additionally, a fragment of the following

nuclear DNA genes was sequenced: exon 1 of rhodopsin, single exon of *Rag1*, and exon 2 of *Cxcr4*, yielding alignments of 315, 555 and 675 bp, respectively. Methods for DNA extraction, amplification (including primers) and sequencing are given elsewhere²². The following primers were developed for this study: Rag1-B, 5'-ATGGGAGATGTGAGTGA RAARCA-3'; Rag1-C, 5'-GGAGATGTTAGTGAGAARCAAYGG-3'; Rag1-D, 5'-GCTGC ATTTCCRATRTACAGTG-3'; Rag1-E, 5'-TCCGCTGCATTTCCRATGTGCA-3'; and Rag1-F, 5'-CCAATGTCGAGTGAARGCRTC-3'. CXCR4-A, 5'-CTGCTGGGCAT CATTGGRAAYGG-3'; CXCR4-B, 5'-ATCATTGGCAATGGAYTRGT-3'; CXCR4-C, 5'-GTCATGGGCTAYCARAAGAA-3'; CXCR4-D, 5'-AGGACAATGACWGAYAAATA-3'; CXCR4-E, 5'-AGGACAATGACWGAYAAAGTA-3'; CXCR4-F, 5'-TTGAATTTGGCCCC AGGAARGC-3'; and CXCR4-G, 5'-AGGCAACAGTGGAARAANGC-3'.

Alignment and phylogenetic analyses

Sequences were aligned using ClustalX 1.64 and ambiguous sections were excluded for subsequent analyses. Phylogenetic analyses were performed using PAUP* 4.0b10 (ref. 23). Plots of transitions and transversions against uncorrected and GTR-corrected pairwise distances indicated that none of the fragments showed saturation. Partition homogeneity tests revealed no significant incongruence among different fragments. Heuristic maximum parsimony searches were executed in 10,000 replicates with all characters unordered and equally weighted. Clade support under maximum parsimony was calculated using decay indices and nonparametric bootstrapping. Appropriate likelihood models were determined using the software Modeltest 3.06 (ref. 24). We also conducted 250 replicated metaGA searches²⁵ using MetaPIGA 1.0.2b, each with strict consensus pruning among four populations, using a HKY + Γ + I model (the most parameter-rich model implemented in MetaPIGA) with the T_i/T_e ratio optimized every 200 generations. The 1,000 resulting trees were used to compute a majority-rule consensus tree and calculate posterior branch support values. Bayesian analyses were performed using the software MrBayes 2.01 (ref. 26) under the GTR + Γ + I model. Four chains were run simultaneously for 1,000,000 generations and trees were sampled every 100 cycles. Likelihood scores reached stationarity well before 100,000 generations, but to be on the safe side we discarded the first 2,000 trees as the 'burn in'. Hence bayesian posterior probabilities were estimated as the 50% majority-rule consensus tree of the 8,000 last sampled trees.

Dating estimates

Mean and 95% credibility interval values for node times were estimated from 1,443 bp of nuclear DNA on the consensus phylogram (Fig. 2) using a bayesian method that allows correlated rate changes between the nodes of a tree⁹. After incorporating in our data set the homologous fragments of the nuclear genes of *Danio rerio*, *Homo sapiens* and *Gallus gallus* (retrieved from GenBank), as well as the caudates *Hynobius formosanus* and *Salamandra salamandra*, we were able to use three calibrations simultaneously: (1) the divergence of amniotes from amphibians at a minimum of 338 Myr ago²⁷; (2) the divergence of diapsids from synapsids at about 310 Myr ago²⁸; and (3) the minimum age of Cryptobranchioidea (164 Myr ago; here represented by the *Hynobius-Salamandra* split)²⁹ (see Supplementary Information).

Received 20 May; accepted 1 September 2003; doi:10.1038/nature02019.

1. AmphibiaWeb [online] (<http://amphibiaweb.org/>) (Berkeley, California, 2003).
2. Frost, D. R. *Amphibian Species of the World: an Online Reference* V2.21 (<http://research.amnh.org/herpetology/amphibia/index.html>) (American Museum of Natural History, 15 July 2002).
3. Ford, L. S. & Cannatella, D. C. The major clades of frogs. *Herpetol. Monogr.* 7, 94–117 (1993).
4. Duellman, W. E. & Trueb, L. *Biology of Amphibians* 1–670 (John Hopkins Univ. Press, Baltimore/London, 1986).
5. Hay, J. M., Ruvinsky, I., Hedges, S. B. & Maxson, L. R. Phylogenetic relationships of amphibian families inferred from DNA sequences of mitochondrial 12S and 16S ribosomal RNA genes. *Mol. Biol. Evol.* 12, 928–937 (1995).
6. Ruvinsky, I. & Maxson, L. R. Phylogenetic relationships among bufonoid frogs (Anura: Neobatrachia) inferred from mitochondrial DNA sequences. *Mol. Phylogenet. Evol.* 5, 533–547 (1996).
7. Feller, A. E. & Hedges, S. B. Molecular evidence for the early history of living amphibians. *Mol. Phylogenet. Evol.* 9, 509–516 (1998).
8. Nussbaum, R. A. Phylogenetic implications of amplexic behavior in sooglossid frogs. *Herpetologica* 36, 1–5 (1980).
9. Thorne, J. L. & Kishino, H. Divergence time and evolutionary rate estimation with multilocus data. *Syst. Biol.* 51, 689–702 (2002).
10. Avise, J. C. & Johns, G. C. Proposal for a standardized temporal scheme of biological classification for extant species. *Proc. Natl Acad. Sci. USA* 96, 7358–7363 (1999).
11. Savage, J. M. in *Evolutionary Biology of the Anurans: Contemporary Research on Major Problems* (ed. Vial, J. L.) 351–445 (Univ. Missouri Press, Columbia, 1973).
12. Coffin, M. F. & Rabinowitz, P. D. in *Geology and Geophysics of Continental Margins* (eds Watkins, J. S., Zhiqiang, F. & McMillen, K. J.) 207–246 (American Association of Petroleum Geologists, Tulsa, Oklahoma, 1992).
13. Tyler, M. J. Phylogenetic significance of the superficial mandibular musculature and vocal sac structure of sooglossid frogs. *Herpetologica* 41, 173–176 (1985).
14. Hay, W. W. et al. in *Evolution of the Cretaceous Ocean-Climate System* (eds Barrera, E. & Johnson, C.) 1–48 (Geological Society of America, Boulder, 1999).
15. Wilkinson, M., Sheps, J. A., Oommen, O. V. & Cohen, B. L. Phylogenetic relationships of Indian caecilians (Amphibia: Gymnophiona) inferred from mitochondrial rRNA gene sequences. *Mol. Phylogenet. Evol.* 23, 401–407 (2002).
16. Bossuyt, F. & Milinkovitch, M. C. Amphibians as indicators of Early Tertiary 'Out-of-India' dispersal of vertebrates. *Science* 292, 92–95 (2001).
17. Spinar, Z. V. & Hodrova, M. New knowledge of the genus *Indobatrachus* (Anura) from the Lower Eocene of India. *Amphib. Reptil.* 6, 363–376 (1985).
18. Lynch, J. D. *Evolutionary Relationships, Osteology and Zoogeography of Leptodactylid Frogs* 1–238 (Univ. Kansas, Museum of Nat. Hist. Misc. Publ. 53, Lawrence, 1971).

19. Nussbaum, R. A. Mitotic chromosomes of Sooglossidae (Amphibia: Anura). *Caryologia* **32**, 279–298 (1979).
20. Krause, D. W., Prasad, G. V. R., von Koenigswald, W., Sahni, A. & Grine, F. E. Cosmopolitanism among Gondwanan Late Cretaceous mammals. *Nature* **390**, 504–507 (1997).
21. Krause, D. W. Fossil molar from a Madagascar marsupial. *Nature* **412**, 497–498 (2001).
22. Bossuyt, F. & Milinkovitch, M. C. Convergent adaptive radiations in Madagascar and Asian ranid frogs reveal covariation between larval and adult traits. *Proc. Natl Acad. Sci. USA* **97**, 6585–6590 (2000).
23. Swofford, D. L. *PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods)* Version 4.0b10 (Sinauer Associates, Sunderland, Massachusetts, 2002).
24. Posada, D. & Crandall, K. A. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818 (1998).
25. Lemmon, A. R. & Milinkovitch, M. C. The Metapopulation genetic algorithm: an efficient solution for the problem of large phylogeny estimation. *Proc. Natl Acad. Sci. USA* **99**, 10516–10521 (2002).
26. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
27. Ruta, M., Coates, M. I. & Quicke, D. L. J. Early tetrapod relationships revisited. *Biol. Rev.* **78**, 251–345 (2003).
28. Benton, M. J. *Vertebrate Palaeontology* 2nd edn, 1–452 (Chapman & Hall, London, 1997).
29. Gao, K. & Shubin, N. H. Earliest known crown-group salamanders. *Nature* **422**, 424–428 (2003).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We are grateful to the Kerala Forest Department for collecting permits (to S.D.B.), and to J. Nagaraju (Molecular Genetics Laboratory, CDFD, Hyderabad) for support in sequencing *Nasikabatrachus*. P. Mardulyn, M. C. Milinkovitch and K. Roelants provided comments on an earlier draft. D. Cannatella, R. Brown, S. Donnellan, R. Drewes, J. Vindum, R. Nussbaum, M. Vences, E. Scott and D. Wake provided tissue samples. A. Mannaert assisted in the molecular work. D. Adriaens, S. Devaere, F. Verschoten and A. Herrel were supportive in osteological analyses. J. Thorne made his software available. M. Coates provided suggestions for calibration.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to F.B. (fbossuyt@vub.ac.be). Sequences have been deposited in GenBank under accession numbers AY364167–AY364226 and AY364340–AY364407.

Females increase offspring heterozygosity and fitness through extra-pair matings

Katharina Foerster¹, Kaspar Delhey¹, Arild Johnsen¹, Jan T. Lifjeld² & Bart Kempnaers¹

¹Max Planck Research Centre for Ornithology, PO Box 1564, D-82305 Starnberg (Seewiesen), Germany

²Zoological Museum, University of Oslo, PO Box 1172 Blindern, N-0318 Oslo, Norway

Females in a variety of species commonly mate with multiple males, and there is evidence that they benefit by producing offspring of higher genetic quality^{1–3}; however, the nature of these genetic benefits is debated^{1–4}. Enhanced offspring survival or quality can result from intrinsic effects of paternal genes—‘good genes’—or from interactions between the maternal and paternal genomes—‘compatible genes’^{1–5}. Evidence for the latter process is accumulating^{2,6}: matings between relatives lead to decreased reproductive success, and the individual level of inbreeding—measured as average heterozygosity—is a strong fitness predictor^{7–13}. Females should thus benefit from mating with genetically dissimilar males^{2,14}. In many birds, social monogamy restricts mate choice, but females may circumvent this by pursuing extra-pair copulations^{15,16}. Here we show that female blue tits, *Parus caeruleus*, increase the heterozygosity of their progeny through extra-pair matings. Females thereby produce offspring of higher reproductive value, because less inbred individuals have increased survival chances, a more elaborate

male secondary sexual trait (crown colour) and higher reproductive success. The cost of inbreeding may therefore be an important factor driving the evolution of female extra-pair mating.

Previous work suggests that blue tit females gain genetic benefits from mating with multiple males^{17,18}: broods of socially monogamous pairs frequently contain extra-pair young, which are in better condition and are more likely to fledge than their within-pair nest mates¹⁸. So far, the genetic mechanism underlying this fitness advantage is unknown. Individual genetic diversity (heterozygosity) reflects the level of inbreeding and influences survival and fitness in various species^{7,9–13}. High individual heterozygosity reduces the likelihood that recessive deleterious alleles are expressed, or increases the number of potentially useful gene products (for example, at major histocompatibility complex (MHC) genes)⁴. Therefore, mating with individuals carrying dissimilar alleles can be advantageous^{2,14}. If restricted social mate choice leads to pairings of genetically similar individuals, extra-pair copulations provide a mechanism to counteract inbreeding depression by increasing the genetic diversity, and hence fitness, of offspring¹⁶. If female blue tits use this strategy, extra-pair young should be more heterozygous than within-pair young and individual heterozygosity should influence fitness.

We examined genetic parentage, breeding success, and offspring and adult survival in a population of individually marked blue tits breeding in nest boxes in the Viennese Forest, Austria (48°13' N, 16°20' E), from 1998 to 2001. Extra-pair young were more heterozygous than their maternal half-siblings sired by the social father (Fig. 1). This suggests that females are less related to extra-pair fathers than to their social mates. We tested this with all known extra-pair fathers but did not find the expected result (social male, mean relatedness $\pm$ standard error of the mean (s.e.m.) = -0.016 ± 0.016 ; extra-pair males, -0.029 ± 0.014 , paired *t*-test, *t* = 0.62, *n* = 96 dyads, *P* = 0.54; 95% confidence interval for the difference in relatedness between female and social/extra-pair male (-0.029 , 0.056)). However, 28% of all extra-pair young were sired by unknown males (not found breeding in the study area). Extra-pair young produced by these non-local fathers mainly accounted for the difference in heterozygosity (Fig. 1). Among known extra-pair fathers, close neighbours did not increase offspring heterozygosity, whereas local non-neighbours tended to sire more heterozygous extra-pair young (Fig. 1). Therefore, only non-neighbouring extra-pair males should be less related to the female than her social partner. We failed to find this, but our sample size of broods with known extra-pair fathers is small (local non-neighbours, *n* = 15; non-local males, *n* = 7, both *P* > 0.7). We then tested whether females were generally less related to more distantly breeding males. We measured the distance and calculated the relatedness between all breeding males and females in each of three years. We then averaged relatedness for each of 14 distance classes, ranging from 0 m (social partner) to 1,300 m. In each year, genetic similarity decreased with breeding distance (1998, *r_p* = -0.48 , *n* = 13, *P* = 0.096; 1999, *r_p* = -0.63 , *n* = 14, *P* = 0.015; 2000, *r_p* = -0.67 , *n* = 13, *P* = 0.017; Fisher's combined probability, *P* = 0.003). Thus, a genetic structure in this population enables females to obtain genetically less similar extra-pair partners by copulating with males that breed further away.

Our data show that females increase their offspring's heterozygosity by extra-pair copulations with non-neighbouring males, which accounts for 50% of all extra-pair young (*n* = 385). Extra-pair matings with close neighbours did not lead to increased offspring heterozygosity, but are nevertheless actively pursued by female blue tits (ref. 17 and our own unpublished data). Local extra-pair males were older, larger and sang longer strophes than the cuckolded males in a Belgian population^{17,18}. Similarly, we found here that extra-pair males were older and larger than the social males if they were close neighbours (age in years: social male, mean $\pm$ s.e.m. = 1.6 ± 0.1 ; extra-pair male, 2.0 ± 0.1 , Wilcoxon's

matched pairs signed ranks test, $Z = -2.93$, $n = 79$, $P = 0.003$; tarsus length in mm: social male, 17.10 ± 0.06 ; extra-pair male, 17.32 ± 0.05 , paired t -test, $t = 2.66$, $n = 73$, $P = 0.009$) but not if they were non-neighbours (age: social male, 1.7 ± 0.1 ; extra-pair male, 1.6 ± 0.1 , $Z = -0.04$, $n = 36$, $P = 0.97$; tarsus length: social male, 17.24 ± 0.08 ; extra-pair male, 17.08 ± 0.08 , $t = -1.31$, $n = 31$, $P = 0.20$). This suggests that females gain different benefits from extra-pair copulations with different males. Copulations with close neighbours may be pursued to produce offspring with 'good genes', whereas copulations with non-local males may be sought to increase offspring heterozygosity. Thus, both the good genes and compatible genes hypothesis may explain why females mate with multiple males in the same population.

Females only benefit from producing heterozygous young if the individual level of inbreeding influences fitness. Fledglings that survived and bred in the study area (local recruits) were more heterozygous than their non-recruiting nestmates (recruits, mean $\pm$ s.e.m. = 1.04 ± 0.02 ; non-recruits, 1.00 ± 0.02 , Wilcoxon's matched pairs signed ranks test, $Z = -2.93$, $n = 52$, $P = 0.003$). Furthermore, individual heterozygosity predicted survival of first-year breeding females to the next season (Table 1). This was not the case in males (Table 1). Selection on males may take place in the first year (for example, during territory establishment in autumn), whereas selection on females may be stronger after breeding. As most local recruits are males (72%, $n = 53$ recruits), the sample size is too small to test for sex-specific effects of heterozygosity on the probability of recruitment.

Individual levels of inbreeding also predicted the annual reproductive output of both sexes. In adult females heterozygosity was positively related to clutch size (Table 1). Male heterozygosity did not affect the total number of eggs sired (Table 1), but it predicted success in raising young: broods of heterozygous males had a higher fledging success than those of less heterozygous males (Table 1), and heterozygous males ultimately produced more local recruits (Table 1). This suggests that heterozygous individuals are of higher quality. If so, heterozygous males may also be able to produce more elaborate secondary sexual traits—as predicted by the hetero-

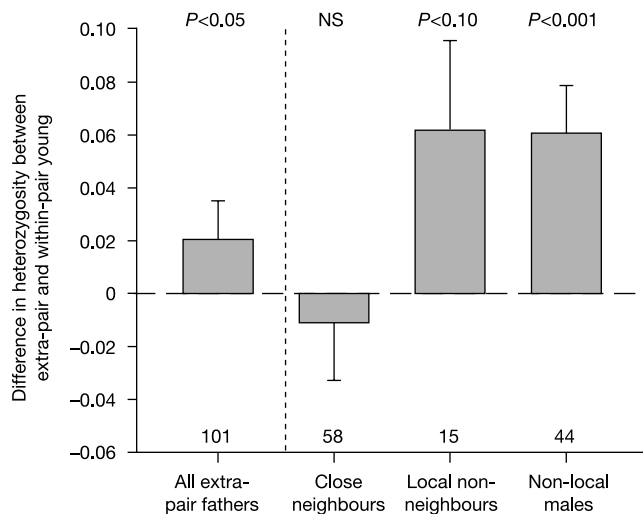


Figure 1 Mean difference ($\pm$ s.e.m.) in standardized heterozygosity between extra-pair young and their within-pair nestmates. All broods with mixed paternity: Wilcoxon's matched pairs signed ranks test, $Z = -2.15$, $P = 0.032$. Broods with extra-pair young fathered by close neighbours (breeding 1–2 territories or 50–200 m from the focal nest), $Z = -0.05$, $P = 0.96$; local non-neighbours (breeding 3–8 territories or 150–800 m away), $Z = -1.82$, $P = 0.069$; or non-local males (not breeding in the study area) $Z = -3.38$, $P = 0.001$. Broods containing extra-pair young from different categories of fathers are included more than once. NS, not significant. Numbers above the x axis indicate numbers of broods studied.

zygosity theory^{4,19}—and therefore have a mating advantage. Several studies suggest that crown coloration is a secondary sexual signal in the blue tit^{20–22}. Accordingly, more heterozygous males had crown feathers with a higher ultraviolet component (chroma) and a more short-wave-shifted hue (Fig. 2).

If females obtain genetic benefits from matings with less-related extra-pair males by producing more successful offspring, they may also prefer these males as social partners¹⁴. However, female relatedness to the social partner (mean $\pm$ s.e.m. = 0.001 ± 0.012) did not differ from the average relatedness of females to all other males present in the same year (-0.003 ± 0.002 ; paired t -test, $t = 0.42$, $n = 219$, $P = 0.68$). Thus, we cannot reject the hypothesis that individuals mated at random with regard to relatedness. This may indicate that the costs of inbreeding avoidance exceed those of

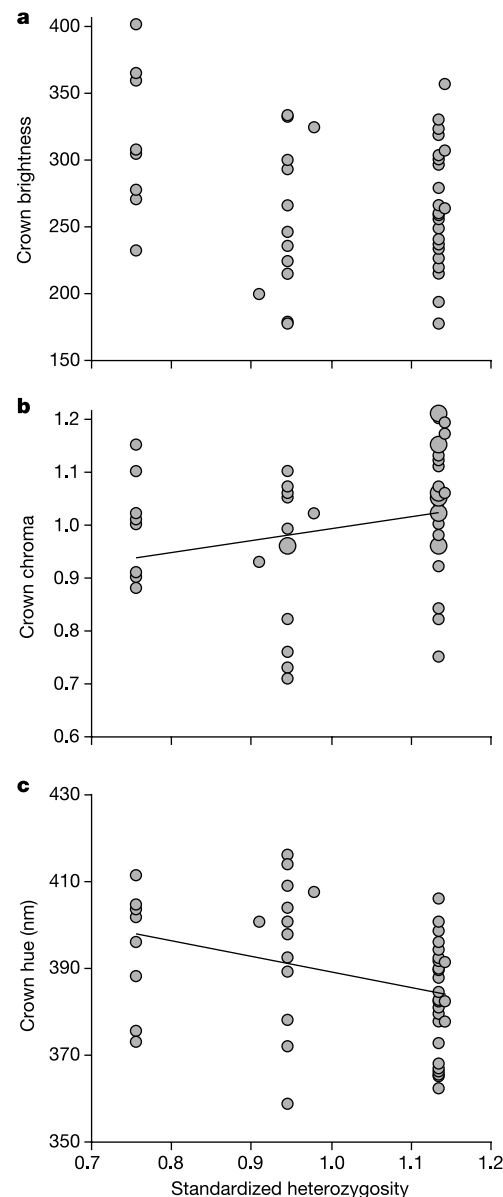


Figure 2 Correlation between standardized heterozygosity and crown coloration in 49 male blue tits. Small and large dots represent one and two data points, respectively. Lines are based on least-squares regression. **a**, Brightness, Spearman rank correlation coefficient $r_s = -0.13$, $P = 0.38$. **b**, Chroma, $r_s = 0.41$, $P = 0.004$ (ultraviolet chroma, $r_s = 0.46$, $P = 0.001$; not shown). **c**, Hue, $r_s = -0.39$, $P = 0.006$. As chroma and hue are highly correlated, we also calculated the first principal component and related this to male heterozygosity ($r_s = 0.46$, $P = 0.001$; not shown).

Table 1 **Effects of standardized heterozygosity on adult survival and reproductive success**

Dependent variable	<i>n</i>	Independent variable	<i>b</i> *	Test statistic	<i>P</i> -value	Explained variance†
Female						
Survival‡	102	Heterozygosity	3.69	$\chi^2_1 = 6.45$	0.011	0.082
Clutch size§	126	Laying date	-0.10	$F_{1,123} = 11.76$	0.001	0.087
		Heterozygosity	2.35	$F_{1,123} = 6.02$	0.016	0.047
Male						
Survival‡	103	Heterozygosity	-0.28	$\chi^2_1 < 0.05$	0.83	—
No. of sired eggs	113	Year	—	$F_{2,106} = 3.64$	0.030	0.064
		Laying date	—	$F_{1,106} = 0.19$	0.67	—
		Year × laying date	—	$F_{2,106} = 3.81$	0.025	0.067
		Heterozygosity	2.03	$F_{1,106} = 0.58$	0.45	—
Young fledged (%)¶	96	Year	—	$F_{2,89} = 3.09$	0.051	0.052
		Laying date	0.12	$F_{1,89} = 4.68$	0.033	0.040
		Year × laying date	—	$F_{2,89} = 3.40$	0.038	0.058
		Heterozygosity	2.08	$F_{1,89} = 5.09$	0.027	0.043
No. of sired recruits#	96	Year	—	$\chi^2_2 = 18.99$	<0.001	0.151
		Heterozygosity	2.96	$\chi^2_1 = 7.86$	0.005	0.063

Age was estimated as the number of seasons individuals were seen in the study area. We assumed that individuals absent in a given year had died, and that individuals first caught after their first year were two years old. Confounding variables with $P > 0.1$ were deleted from the final model, unless a first-order interaction was significant.

* Unstandardized regression coefficient.

† Proportion of the total variance explained by the independent variable.

‡ Survival of one-year-old birds. Binary logistic regression including year in the full model.

§ Analysis of co-variance (ANCOVA) including year, age, laying date and first-order interactions in the full model.

|| Within-pair and extra-pair combined. ANCOVA including year, age, laying date and first-order interactions in the full model.

¶ GLM with binomial error including year, age, laying date and first-order interactions in the full model. Scale parameter set to 4.427.

Sum of within- and extra-pair offspring that bred in the study area the following year. GLM with Poisson error including year, age, laying date and first-order interactions in the full model.

inbreeding, or it may reflect a lack of mate choice opportunities at the time of pair formation. In our population, most pairs stay together year-round, and territory vacancies caused by death of one of the partners are filled as they arise throughout winter and spring. Thus, only a limited selection of potential social mates is available at any one time. Extra-pair matings may then provide the only opportunity to obtain more heterozygous offspring¹⁶.

Should females prefer heterozygous males as partners? Mating with such males increases the likelihood of producing heterozygous offspring, because heterozygous individuals carry more rare alleles^{4,23}. However, these alleles are likely to be deleterious recessives and will therefore not increase the fitness of their bearer². Females would only benefit if males are heterozygous at loci showing a fitness advantage of newly evolved, rare alleles (for example, MHC loci)². Thus, if heterozygosity at microsatellite loci relates to heterozygosity at such coding loci, heterozygous males might be more successful in obtaining paternity. However, males that lost paternity did not differ in heterozygosity from non-cuckolded males (mean $\pm$ s.e.m. = 1.01 ± 0.02 and 1.01 ± 0.02 , respectively, Mann–Whitney U -test, $U = 1,624.0$, $n_1 = 57$, $n_2 = 60$, $P = 0.63$). Similarly, males gaining extra-pair paternity did not differ from other males (0.99 ± 0.03 and 1.02 ± 0.02 , respectively, $U = 1,328.0$, $n_1 = 36$, $n_2 = 81$, $P = 0.44$), nor did social males (1.01 ± 0.01) differ from their cuckolders (0.98 ± 0.02 , Wilcoxon's matched pairs signed ranks test, $Z = -1.60$, $n = 115$, $P = 0.11$). Hence, there is no evidence that female blue tits preferred heterozygous males as sires for their offspring. However, females may derive non-genetic benefits from heterozygous social partners. If less inbred males are healthier or more viable, they may provide better resources (for example, territory, parental care). Our data support this, as heterozygous males were more successful in raising young (Table 1). As crown coloration correlates with individual heterozygosity, females might prefer colourful males as social mates. Indeed, ultraviolet crown coloration seems to influence social mating decisions (assortative mating²⁰) and offspring sex allocation²² in this species.

We have shown that the individual level of inbreeding strongly predicts fitness in a natural population of blue tits, and that females gain genetic benefits from a commonly observed mating behaviour (extra-pair copulations) by producing more heterozygous offspring. This supports the idea that females mate multiply to obtain compatible genes for their offspring, thereby avoiding the costs of inbreeding^{2–6,16,24}. This could be an important factor in the evolution of female promiscuity. □

Methods

Field methods

For blood sampling (20–100 μ l) we caught all adults in the study area before breeding or during chick feeding. Individuals were aged (one year or older) and sexed on the basis of plumage characteristics. We recorded social pairing and breeding success during frequent visits to all nest boxes, and by regular behavioural observations. We collected all unhatched eggs and dead nestlings, and bled surviving nestlings before fledging to obtain DNA. We banded all 14-day-old nestlings and determined fledgling survival and recruitment by recapture from December onwards.

Reflectance spectrometry

For 49 males (caught when feeding young in May 2001), we measured the reflectance of the crown feathers using a S2000 spectrometer with illumination by a DH-2000 deuterium-halogen lamp (Ocean Optics). For each individual, we averaged five scans from different spots on the crown. The probe, mounted inside a black plastic tube, was held perpendicular to the crown plumage ($0^\circ/0^\circ$, illumination and recording). Using SpectraWin 3.1, spectra were recorded in 0.38-nm steps from 320–700 nm and expressed relative to a WS-2 white reflection standard (TOP Sensor Systems). Spectra were smoothed by a running average over 40 recording steps. We calculated parameters of the three dimensions of colour perception: brightness (total reflectance, $R_{320-700}$), chroma (spectral purity, $(R_{\max} - R_{\min})/R_{\text{average}}$) and hue (wavelength of peak reflectance, $\lambda(R_{\max})$)²⁰. We also calculated ultraviolet chroma, $(R_{320-400}/R_{320-700})$, which addresses the specific importance of this wavelength range, to allow direct comparison with previous studies^{20,22}.

Microsatellite paternity analysis

We genotyped 2,452 offspring (embryos or nestlings; 71% of all offspring (eggs) produced in the study area) of 248 broods using five (1998) or six (1999–2001) microsatellite markers (*Pca3*, *Pca4*, *Pca7*, *Pca9* (ref. 25); *Poccl*, *Poccl6* (ref. 26); *Phtr3* (ref. 27)) with a combined exclusion probability >0.995 (1998) or >0.999 (1999–2001). We sampled all males that bred in the nest boxes, except six that we could not catch. Only five pairs bred in natural cavities, but the males were caught in winter and therefore included in the paternity analysis. We excluded the putative father if two or more loci showed non-matching alleles with the offspring. We assumed that only one mismatch with the social father (after repeated genotyping) was due to mutation and assigned these chicks to the putative father (probability of false inclusion, PFI²⁸, excluding the mismatching locus: 2.02×10^{-5} to 4.57×10^{-3} , $n = 11$).

For 274 of 385 extra-pair young we found fathers with completely matching genotype. When two males matched the genotype of an extra-pair chick ($n = 4$), we assigned paternity to the male who fathered other offspring in the nest. For eight extra-pair young, we found putative fathers with only one mismatching allele. If the male sired other young in the same nest, we assumed that the mismatches were caused by mutations and accepted the male as the genetic father ($n = 2$, PFI $< 8.48 \times 10^{-5}$). We were not able to assign the genetic father to 109 extra-pair young. As we genotyped almost all males breeding in the study area, we assumed that these young were sired by non-local males (either breeding outside the study area or floaters). Seven non-local males observed in the area during winter sired extra-pair young.

Individual genetic diversity and pairwise relatedness

We assessed individual genetic diversity at 5–7 microsatellite loci (see above) with observed heterozygosity ranging from 0.83 to 0.93. For females and offspring *Phtr3* was excluded, because this locus is hemizygous in females. Individual heterozygosity (number of heterozygous loci/total number of typed loci) varied between 0.6 and 1.0 in adults and

between 0.2 and 1.0 in nestlings. As not all individuals were typed with the same set of microsatellite markers, we calculated standardized individual heterozygosity¹⁰ (proportion of heterozygous loci/mean heterozygosity of typed loci). This measure is used in all analyses and is referred to as 'individual heterozygosity' for simplicity. We also calculated two other, previously suggested measures of individual genetic diversity: internal relatedness¹⁰ and mean d^2 -value²⁹ where d^2 is the squared length difference between the alleles at a locus. Heterozygosity, standardized heterozygosity and internal relatedness were highly correlated (all $r > 0.89$, $P < 0.0001$) and all three measures led to similar conclusions (data not shown). Mean d^2 -values were weakly correlated with the other measures ($r = 0.24$ – 0.32 , all $P < 0.0001$) and explained comparatively little variation in individual fitness.

The assumption that heterozygosity measured at a limited number of loci reflects genome-wide heterozygosity can be questioned³⁰. Any relationship between average heterozygosity and fitness might be due to close linkage between a microsatellite locus and loci coding for fitness-related traits. To check whether the effect of overall heterozygosity on fitness depended on one or a few of the microsatellite loci, we repeated all analyses for each locus separately (data not shown). Heterozygosity at six out of seven loci significantly affected at least one fitness measure, and different fitness measures were predicted by heterozygosity at different loci, so we only present data on overall heterozygosity.

We calculated pairwise relatedness between partners using Relatedness 5.0 (<http://www.gsfn.net.us/GSoft.html>). The average relatedness between adult males and females (expected $r = 0$) in our study population was -0.0007 (s.e.m. calculated by 'jack-knifing' over loci = 0.0009). In a random sample of 100 broods, the average relatedness among full siblings (expected $r = 0.5$) was 0.47 (s.e.m. = 0.02). The mean inbreeding coefficient of our population (F_{is}), calculated from the microsatellite data using Fstat V2.9.3, equals -0.006 . The observed level of inbreeding was 3.3‰: two of 61 local recruits (with known parents) mated with close relatives ($r = 0.5$ and 0.125).

Received 21 March; accepted 25 July 2003; doi:10.1038/nature01969.

- Jennions, M. D. & Petrie, M. Why do females mate multiply? A review of the genetic benefits. *Biol. Rev.* **75**, 21–64 (2000).
- Tregenza, T. & Wedell, N. Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol. Ecol.* **9**, 1013–1027 (2000).
- Zeh, J. A. & Zeh, D. W. Reproductive mode and the genetic benefits of polyandry. *Anim. Behav.* **61**, 1051–1063 (2001).
- Brown, J. L. A theory of mate choice based on heterozygosity. *Behav. Ecol.* **8**, 60–65 (1997).
- Zeh, J. A. & Zeh, D. W. The evolution of polyandry I: intragenomic conflict and genetic incompatibility. *Proc. R. Soc. Lond. B* **263**, 1711–1717 (1996).
- Tregenza, T. & Wedell, N. Polyandrous females avoid costs of inbreeding. *Nature* **415**, 71–73 (2002).
- Thornhill, N. W. *The Natural History of Inbreeding and Outbreeding: Theoretical and Empirical Perspectives* (Univ. Chicago Press, Chicago, 1993).
- Kempnaers, B., Adriaens, E., van Noordwijk, A. J. & Dhondt, A. A. Genetic similarity, inbreeding and hatching failure in blue tits: are unhatched eggs infertile? *Proc. R. Soc. Lond. B* **263**, 179–185 (1996).
- Coltman, D. W., Bowen, W. D. & Wright, J. M. Birth weight and neonatal survival of harbour seal pups are positively correlated with genetic variation measured by microsatellites. *Proc. R. Soc. Lond. B* **265**, 803–809 (1998).
- Amos, W. *et al.* The influence of parental relatedness on reproductive success. *Proc. R. Soc. Lond. B* **268**, 2021–2027 (2001).
- Hansson, B., Bensch, S., Hasselquist, D. & Åkesson, M. Microsatellite diversity predicts recruitment of sibling great reed warblers. *Proc. R. Soc. Lond. B* **268**, 1287–1291 (2001).
- Höglund, J. *et al.* Inbreeding depression and male fitness in black grouse. *Proc. R. Soc. Lond. B* **269**, 711–715 (2002).
- Hansson, B. & Westerberg, L. On the correlation between heterozygosity and fitness in natural populations. *Mol. Ecol.* **11**, 2467–2474 (2002).
- Pusey, A. & Wolf, M. Inbreeding avoidance in animals. *Trends Ecol. Evol.* **11**, 201–206 (1996).
- Petrie, M. & Kempnaers, B. Extra-pair paternity in birds: explaining variation between species and populations. *Trends Ecol. Evol.* **13**, 52–58 (1998).
- Blomqvist, D. *et al.* Genetic similarity between mates and extra-pair parentage in three species of shorebirds. *Nature* **419**, 613–615 (2002).
- Kempnaers, B. *et al.* Extra-pair paternity results from female preference for high-quality males in the blue tit. *Nature* **357**, 494–496 (1992).
- Kempnaers, B., Verheyen, G. R. & Dhondt, A. A. Extrajoint paternity in the blue tit (*Parus caeruleus*): female choice, male characteristics, and offspring performance. *Behav. Ecol.* **8**, 481–492 (1997).
- Aparicio, J. M., Cordero, P. J. & Veiga, J. P. A test of the hypothesis of mate choice based on heterozygosity in the spotless starling. *Anim. Behav.* **62**, 1001–1006 (2001).
- Andersson, S., Örnberg, J. & Andersson, M. Ultraviolet sexual dimorphism and assortative mating in blue tits. *Proc. R. Soc. Lond. B* **265**, 445–450 (1998).
- Hunt, S., Cuthill, I. C., Bennett, A. T. D. & Griffiths, R. Preferences for ultraviolet partners in the blue tit. *Anim. Behav.* **58**, 809–815 (1999).
- Sheldon, B., Andersson, S., Griffith, S. C., Örnberg, J. & Sendecka, J. Ultraviolet colour variation influences blue tit sex ratios. *Nature* **402**, 874–877 (1999).
- Mitton, J. B., Schuster, W. S. F., Cothran, E. G. & De Fries, J. C. Correlation between the individual heterozygosity of parents and their offspring. *Heredity* **71**, 59–63 (1993).
- Stockley, P., Searle, J. B., MacDonald, D. W. & Jones, C. S. Female multiple mating behaviour in the common shrew as a strategy to reduce inbreeding. *Proc. R. Soc. Lond. B* **254**, 173–179 (1993).
- Dawson, D. A., Hanotte, O., Greig, C., Stewart, I. R. K. & Burke, T. Polymorphic microsatellites in the blue tit *Parus caeruleus* and their cross-species utility in 20 songbird families. *Mol. Ecol.* **9**, 1941–1944 (2000).
- Bensch, S., Price, T. & Kohn, J. Isolation and characterization of microsatellite loci in a *Phylloscopus* warbler. *Mol. Ecol.* **6**, 91–92 (1997).
- Fridolfsson, A. K., Gyllenstein, U. B. & Jakobsson, S. Microsatellite markers for paternity testing in the willow warbler *Phylloscopus trochilus*: high frequency of extra-pair young in an island population. *Heredity* **126**, 127–132 (1997).
- Jamieson, A. The effectiveness of using co-dominant polymorphic allelic series for (1) checking

pedigrees and (2) distinguishing full-sib pair members. *Anim. Genet.* **25**, 37–44 (1994).

- Pemberton, J. M., Coltman, D. W., Coulson, T. N. & Slate, J. in *Microsatellites, Evolution and Applications* (eds Goldstein, D. B. & Schlötterer, C.) 151–164 (Oxford Univ. Press, Oxford, 1999).
- Turelli, M. & Ginzburg, L. R. Should individual fitness increase with heterozygosity? *Genetics* **104**, 191–209 (1983).

Acknowledgements We thank D. Blomqvist, S. Griffith, D. Hasselquist, L. Keller, M. Milinski, A. Peters, B. Sheldon, C. Wedekind and D. Zeh for comments on the manuscript; K. Carter, D. Kaulfuss, H. Kunc, K. Peer, A. Pösel and A. Türk for help with field and laboratory work; S. Andersson for computing the colour variables; and H. Winkler (Konrad Lorenz Institute for Comparative Ethology) and R.-T. Klumpp and A. Fojt (Institute of Silviculture, University of Agricultural Sciences, Vienna) for logistic support.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.K. (b.kempnaers@erl.ornithol.mpg.de).

RNA molecules stimulate prion protein conversion

Nathan R. Deleault, Ralf W. Lucassen & Surachai Supattapone

Department of Biochemistry, 7200 Vail Building, Dartmouth Medical School, Hanover, New Hampshire 03755, USA

Much evidence supports the hypothesis that the infectious agents of prion diseases are devoid of nucleic acid, and instead are composed of a specific infectious protein¹. This protein, PrP^{Sc}, seems to be generated by template-induced conformational change of a normally expressed glycoprotein, PrP^C (ref. 2). Although numerous studies have established the conversion of PrP^C to PrP^{Sc} as the central pathogenic event of prion disease, it is unknown whether cellular factors other than PrP^C might be required to stimulate efficient PrP^{Sc} production. We investigated the biochemical amplification of protease-resistant PrP^{Sc}-like protein (PrPres) using a modified version³ of the protein-misfolding cyclic amplification method⁴. Here we report that stoichiometric transformation of PrP^C to PrPres *in vitro* requires specific RNA molecules. Notably, whereas mammalian RNA preparations stimulate *in vitro* amplification of PrPres, RNA preparations from invertebrate species do not. Our findings suggest that host-encoded stimulatory RNA molecules may have a role in the pathogenesis of prion disease. They also provide a practical approach to improve the sensitivity of diagnostic techniques based on PrPres amplification.

We previously showed that PrPres amplification *in vitro* shares many specific features with the pathogenic process of prion propagation *in vivo*, including strain and species specificity³. In a typical amplification reaction, diluted prion-infected brain homogenate (0.1% w/v) is mixed with either 5% (w/v) normal brain homogenate or buffer control and incubated overnight at 37 °C. Hamster Sc237 PrPres is amplified about sixfold under these conditions (Fig. 1a, compare M with Sc). While characterizing the biochemical requirements of PrPres amplification reactions, we were surprised to discover that treatment of such reactions with DNase-free, heterogeneous pancreatic RNase abolished PrPres amplification in a dose-dependent manner (Fig. 1a, first panel). *In vitro* PrPres amplification was also abolished by treatment with purified pancreatic RNase A, RNase T1, micrococcal nuclease, or benzonase (Fig. 1a). In a control experiment, we found that addition of RNase A for 1 h after overnight incubation did not reduce the recovery of PrPres already amplified (Supplementary Fig. S1).

By contrast, PrPres amplification was not affected by addition of

RNase V1, which degrades only double-stranded (ds)RNA molecules⁵, or RNase H, which specifically cleaves RNA:DNA hybrids⁶ (Fig. 1b, first and second panels). Taken together, these results suggest that single-stranded (ss)RNA is required for PrPres amplification *in vitro*, but that dsRNA and RNA:DNA hybrids are not. Addition of DNase or the restriction enzyme *EcoRI* did not decrease PrPres amplification, showing that DNA is not required for the process (Fig. 1b, third and fourth panels). Addition of the enzymes apyrase and heparinase III also had no effect on PrPres amplification, suggesting that neither high-energy nucleotides nor molecules containing heparan sulphate are required for PrPres amplification *in vitro* (Fig. 1b, fifth and sixth panels). In control experiments, we confirmed the degradation of the target molecules in each of these reaction mixtures by using appropriate analytical assays for these structures (Supplementary Fig. S2).

To ensure that the commercial nuclease preparations we used were not contaminated with proteases, we measured the levels of PrP^C and PrPres after overnight incubation with various nucleases. These measurements confirmed that levels of PrP^C (Fig. 2a) and input PrPres (Fig. 2b) were both unperturbed by addition of enzymes that inhibited PrPres amplification.

As a control to confirm that abolition of PrPres amplification depends on the stimulatory activity of each inhibitory nuclease, we

added benzonase, micrococcal nuclease and RNase A to PrPres amplification reactions in enzymatically inactive states. Both benzonase and micrococcal nuclease require divalent cations for enzymatic activity, so we inactivated these nucleases by addition of 5 mM EDTA. The active site of RNase A contains a critical histidine residue that can be covalently modified by diethyl pyrocarbonate (DEPC). Therefore, we pre-treated RNase A with DEPC to inhibit its RNase activity, and removed excess DEPC by dialysis. Our results show that none of the three nucleases inhibits PrPres amplification in their inactive states, supporting the hypothesis that intact RNA molecules stimulate this process (Fig. 2c).

To test whether inhibition of PrPres amplification might be mediated by end products of RNase digestion, we measured directly the effect of cyclic 2',3'-guanine monophosphate (GMP) and 3'-cytidine monophosphate (CMP) on PrPres amplification. Neither of these nucleotides inhibited PrPres amplification *in vitro* at concentrations up to 1 mM (Fig. 2d). Our control experiments rule out the possibility that contaminating proteases, steric hindrance, or digestion end-products account for the inhibition of

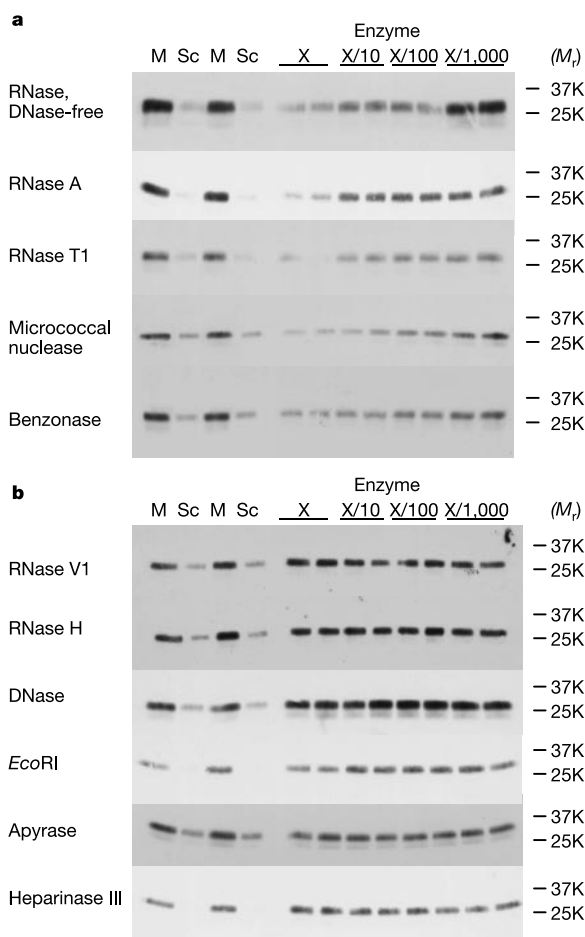


Figure 1 Effect of various enzymes on PrPres amplification. Immunoblots of PrPres amplification reactions. Samples include mixtures of normal and diluted scrapie brain homogenate (M), diluted scrapie brain homogenate control (Sc), and mixtures of normal and diluted scrapie brain homogenate incubated with various enzymes in a dilution series. The values of X (manufacturer's $\text{U } \mu\text{l}^{-1}$) for enzymes are: DNase-free RNase = 0.5; RNase A = 0.005; RNase T1 = 50; micrococcal nuclease = 0.005; benzonase = 2.5; RNase V1 = 0.01; RNase H = 1; RNase-free DNase = 1; *EcoRI* = 1; apyrase = 0.1; heparinase III = 0.025.

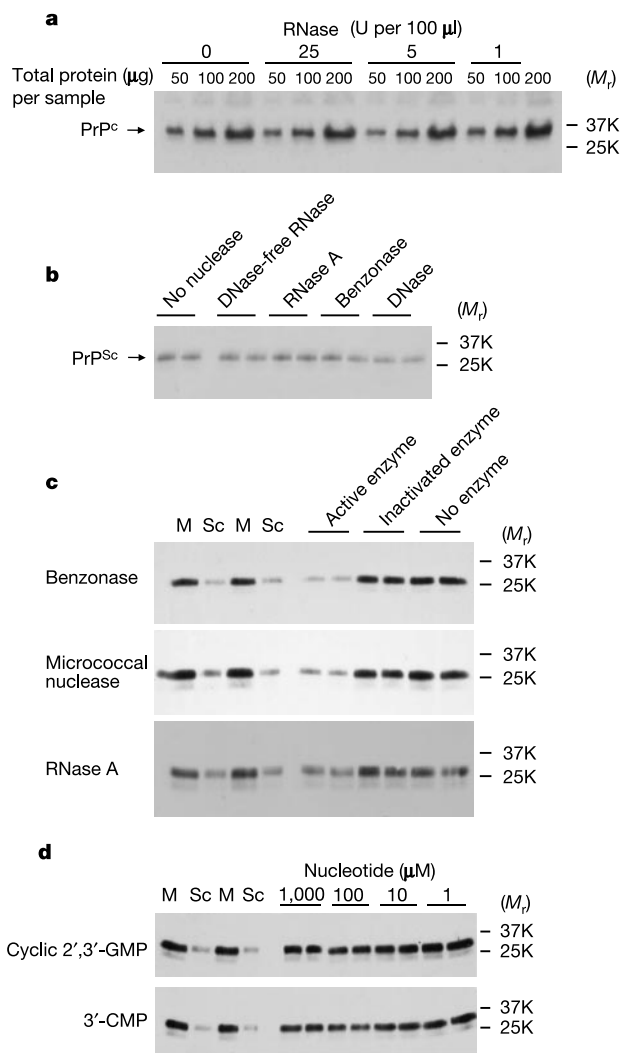


Figure 2 Nucleases do not cause proteolytic, steric or end-product inhibition of PrPres amplification. **a**, Effect of DNase-free RNase on PrP^C levels. Standard amplification mixtures were incubated overnight with RNase. Serial dilutions of each sample are shown. **b**, Effect of nucleases on PrPres levels. Samples of diluted scrapie brain homogenate were treated with specified nucleases at concentrations designated 'X' in Fig. 1. **c**, Effect of inactivated nucleases on PrPres amplification. Samples include mixtures of normal and diluted scrapie brain homogenate (M) and diluted scrapie brain homogenate control (Sc). **d**, Effect of nucleotides on PrPres amplification.

PrPres amplification by specific nucleases. Taken together, these experiments indicate that RNA is required for PrPres amplification *in vitro*.

We next sought to determine whether a preparation of isolated RNA molecules could reconstitute the ability of nuclease-treated normal brain homogenate to amplify PrPres. Remarkably, total RNA isolated from hamster brain successfully reconstituted the ability of benzonase-pre-treated brain homogenate to amplify PrPres in a dose-dependent manner (Fig. 3a). By contrast, purified heparan sulphate proteoglycan (HSPG) failed to reconstitute PrPres amplification (Fig. 3a). Other polyanions, such as ssDNA (Fig. 3b), polyadenylic acid, heparan sulphate, pentosan sulphate and polyglutamic acid (data not shown) also failed to stimulate PrPres amplification. In this and other reconstitution experiments, benzonase-treated control lanes have a greater level of PrPres amplification than the diluted scrapie brain homogenate control samples, indicating that the benzonase pre-treatment reactions were incomplete. Empirically, we found that it was necessary to perform benzonase pre-treatment reactions at 4°C to avoid denaturing PrP^C before the addition of polyanions.

To estimate the molecular size of the RNA species capable of reconstituting PrPres amplification, we fractionated our preparation of total hamster brain RNA by ultrafiltration through a filter with a relative molecular mass cutoff of approximately 100,000 (M_r 100K). Using agarose gel electrophoresis, we detected all of the ribosomal RNA bands in the retentate and all of the transfer RNA in the filtrate (data not shown). Using these samples, we discovered that the filter retentate was capable of reconstituting PrPres amplification to a level slightly lower than unfractionated total brain RNA. By contrast, the filtrate was not able to reconstitute PrPres amplification (Fig. 3c). These data indicate that most of the reconstitution activity is conferred by RNA molecules >100K in size (>300 nucleotides).

There is currently a need to develop more sensitive diagnostic tests for prion disease—this might be achieved by increasing the efficiency of PrPres amplification techniques. We therefore investigated whether addition of total hamster brain RNA could increase the efficiency of PrPres amplification *in vitro* in brain samples not

pre-treated with nuclease. We mixed a more dilute homogenate of prion-infected brain (0.02% w/v) with 5% (w/v) normal brain homogenate overnight without sonication, and measured PrPres amplification. Our results show that addition of total hamster brain RNA to this mixture of intact brain homogenates significantly stimulates PrPres amplification over baseline (Fig. 4a). As a control, we confirmed that addition of RNA did not alter the level of input PrPres or PrP^C in these samples (Supplementary Fig. S3). Densitometric measurements indicate that PrPres in 0.02% (w/v) prion-infected brain homogenate samples is amplified about sixfold after overnight incubation, similar to the PrPres amplification level previously reported for 0.1% (w/v) prion-infected brain homogenate³. By contrast, PrPres in samples amplified with RNA is amplified about 24-fold, indicating that addition of RNA increases the efficiency of *in vitro* PrPres amplification about fourfold. Addition of RNA also increased the efficiency of PrPres amplification of sonicated protein-misfolding cyclic amplification (PMCA) reactions (Supplementary Fig. S4).

To assess the specificity of RNA-mediated stimulation of PrPres amplification, we isolated total RNA from several sources, including *Escherichia coli*, *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, *Drosophila melanogaster*, and mouse and hamster brain. Agarose gel electrophoresis analysis of these preparations revealed the expected band patterns for each species and confirmed that each preparation

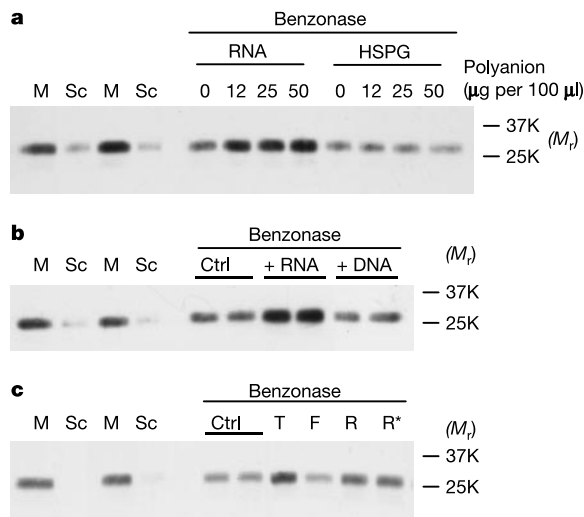


Figure 3 Reconstitution of PrPres amplification with RNA. Immunoblots of PrPres amplification reactions. Samples include mixtures of normal and diluted scrapie brain homogenate (M) and diluted scrapie brain homogenate control (Sc). Indicated samples were pre-treated with benzonase before reconstitution assays as described in Methods. **a**, Reconstitution with total hamster brain RNA or HSPG. **b**, Reconstitution with 0.5 mg ml⁻¹ total hamster brain RNA or 0.5 mg ml⁻¹ random, synthetic 23-base DNA oligonucleotide. **c**, Reconstitution with 0.5 mg ml⁻¹ total (T), filtrate (F), retentate (R) and 50% formamide retentate (R*) samples from RNA fractionated by ultrafiltration.

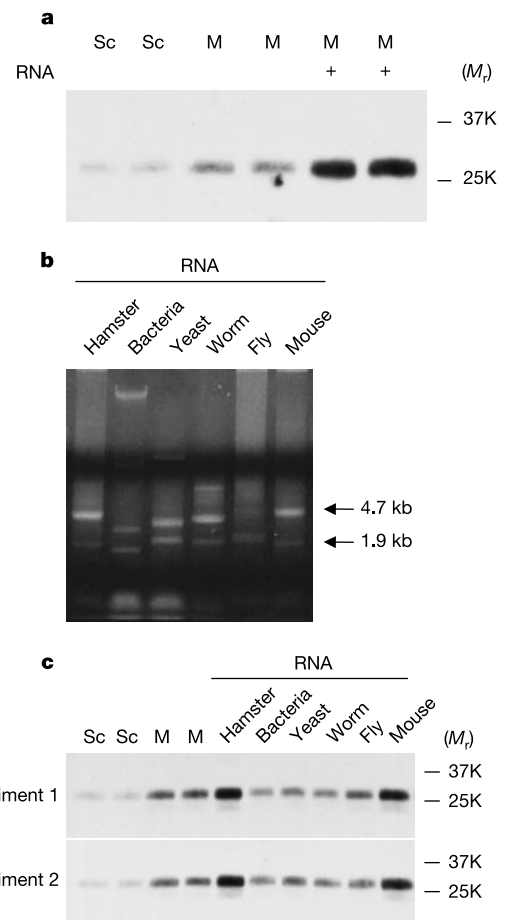


Figure 4 Stimulation of PrPres amplification with RNA. **a**, Immunoblot of PrPres amplification reactions. Samples include mixtures of 5% (w/v) normal and 0.02% (w/v) scrapie brain homogenate (M) and diluted scrapie brain homogenate control (Sc). Indicated samples contained 0.5 mg ml⁻¹ total hamster brain RNA (+). **b**, Agarose gel electrophoresis of total RNA prepared from various species. **c**, Immunoblot of species-specific stimulation of PrPres amplification with RNA. Total RNA (0.5 mg ml⁻¹) prepared from various species was added to PrPres amplification reactions.

contained high-quality, non-degraded RNA (Fig. 4b). Furthermore, each of these preparations was substantially free from contaminants as judged by optical spectroscopy ($A_{260}/A_{280} > 1.9$; where A indicates absorbance and subscript numbers indicate wavelength). Notably, among the six preparations of RNA tested, only hamster and mouse brain RNA could stimulate PrPres amplification *in vitro* (Fig. 4c). This apparent species specificity cannot be attributed to tissue specificity because total hamster liver RNA also stimulated PrPres amplification (data not shown). This argues that mice and hamsters express specific RNA molecules required for PrPres amplification. Additional experiments show that the RNA stimulation activity within the Trizol-extracted hamster brain RNA preparation was irreversibly destroyed by glyoxylation, but not by deproteination, heating to 60 °C, or transient exposure to 50% formamide for 1 h (Supplementary Fig. S5).

If PrPres amplification studies accurately model PrP^{Sc} formation *in vivo*, the results presented here represent a significant advance in our understanding of the mechanism of prion conversion. Previously, it has been shown that purified PrP^C can be converted into protease-resistant PrPres *in vitro* in the absence of cellular cofactors⁷. However, the fact that a 50-fold molar excess of purified PrPres is required to drive conversion of purified PrP^C suggests that efficient PrPres formation may depend on the presence of cellular factors other than PrP^C (ref. 8). On the basis of the results presented here, we propose the hypothesis that specific RNA molecules are cellular cofactors for PrP^{Sc} formation. Consistent with our hypothesis that specific RNA-converting factors stimulate PrP^{Sc} formation, nucleic acids bind avidly to and promote conformational change of recombinant PrP (refs 9–14). However, it is important to note that full-length, refolded recombinant PrP lacking post-translational modifications cannot undergo stoichiometric conversion to PrPres (Supplementary Fig. S6), and therefore the results of biophysical studies using recombinant PrP cannot be directly related to the results described here. It has been proposed that PrP^{Sc} molecules might bind to specific host RNA molecules to generate strain diversity¹⁵. Whether the RNA-converting factors we describe are also involved in generating strain diversity remains to be determined. Finally, it is important to emphasize that the existence of RNA-converting factors is fully consistent with the protein-only hypothesis proposed previously¹, because the nucleic acids we describe are host-encoded and not contained within the infectious agent. □

Methods

Animal and reagent sources

Specific-pathogen-free female golden Syrian hamsters at 3 weeks old were purchased from Charles River Laboratories. Apyrase, DEPC, cyclic 2',3'-GMP, 3'-CMP, heparinase III, heparan sulphate proteoglycan ($M_r > 200K$), polyadenylic acid (M_r 200–2,000K) and polyglutamic acid (M_r 50–100K) were obtained from Sigma; RNase-free DNase, micrococcal nuclease, RNase A and DNase-free RNase were obtained from Roche; RNase T1 was obtained from Epicentre; recombinant benzonase nuclease was purchased from Novagen; EcoRI was obtained from Gibco BRL; and RNase H and RNase V1 were obtained from Ambion.

In vitro PrPres amplification

In vitro PrPres amplification⁸ and PMCA⁴ were performed as previously described, except that normal brain homogenates were prepared with EDTA-free complete protease inhibitors (Roche) to facilitate experiments involving metal-dependent enzymes. Two millimolar MgCl₂ was added to reactions with benzonase and 2 mM CaCl₂ was added to reactions with micrococcal nuclease and apyrase. All amplification and control reactions were performed at 37 °C for 16 h. For PrPres detection, protease digestion was performed with 50 µg ml⁻¹ proteinase K for 1 h at 37 °C and immunoblotting was performed with 3F4 monoclonal antibody (Signet). For PrP^C detection, samples were not subjected to proteinase K digestion before immunoblotting. All protein electrophoresis experiments shown were performed on 12% SDS polyacrylamide gels and reference M_r for such experiments are shown.

Nuclease inactivation

Micrococcal nuclease and benzonase were inhibited with 5 mM EDTA. RNase A (50 µg) was incubated with 1% DEPC in 100 µl at 25 °C for 2 h. After incubation, the reaction was dialysed twice against 1 l 10 mM Tris pH 7.2 at 4 °C using a Pierce 3500 MW Slide-A-Lyzer

minidialysis unit to remove free DEPC. Control samples containing active RNase A were dialysed in parallel. Protein recovery >90% was confirmed by BCA assay (Pierce). Active and inactivated nucleases were added to amplification reactions at concentrations designated 'X' in Fig. 1. 'No enzyme' control samples were processed in parallel.

Reconstitution assays

Nuclease digestion before reconstitution was performed by incubating a batch of normal brain homogenate (10% w/v) with benzonase (final concentration of 2.5 U µl⁻¹) and 2 mM MgCl₂ for 16 h at 4 °C in the absence of detergents. Benzonase was then inactivated by the addition of 5 mM EDTA before reconstitution with RNA or other polyanions.

Preparation and measurement of RNA

RNA was isolated from animals <5 min after death using rotor–stator homogenization, extraction with Trizol reagent (Invitrogen) for 5 min at 25 °C, and isopropanol precipitation according to manufacturer's instructions, using RNase-free reagents, containers and equipment. For yeast, cell walls were disrupted during extraction as previously described, using Trizol in place of phenol¹⁶. All RNA solutions were alcohol precipitated, washed and resuspended in RNase-free water before use. The concentration and purity of each solution was determined by spectroscopic measurement of absorbance at $\lambda_1/\lambda_2 = 260/280$ nm and confirmed by electrophoresis on 1% agarose gels stained with ethidium bromide.

RNA size fractionation

Total hamster brain RNA (0.4 mg) was diluted into 0.8 ml RNase-free water, loaded in 0.2-ml batches onto four separate Schleicher and Schuell Centrex UF-05 (100K cutoff) ultrafiltration devices, and centrifuged for 15 min at 3,000g. The devices were then washed with an equal volume of water. The filtrates were pooled and retentate fractions collected by briefly centrifuging the ultrafiltration devices upside down into new microcentrifuge tubes. Parallel samples of denatured retentate were prepared in 50% formamide to disrupt all intra- and intermolecular interactions.

Reverse transcriptase polymerase chain reaction

RT-PCR was performed using the One Step RNA PCR kit (AMV) from Takara/Fisher following the manufacturer's instructions, using the PrP-specific primers 5'-CGAACC TTGGCTACTGGCTGCTG-3' and 5'-GCTTGATGGTGATATTGACGCGATC-3', and the following parameters: reverse transcription at 50 °C for 15 min, heat inactivation of reverse transcriptase at 94 °C for 2 min, ×25 PCR cycles (94 °C for 30 s, 55 °C for 30 s, 72 °C for 90 s). Products were run on a 1% agarose gel and stained with ethidium bromide.

Received 4 July; accepted 31 July 2003; doi:10.1038/nature01979.

1. Prusiner, S. B. Novel proteinaceous infectious particles cause scrapie. *Science* **216**, 136–144 (1982).
2. Prusiner, S. B. (ed.) *Prion Biology and Diseases* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York, 1999).
3. Lucassen, R., Nishina, K. & Supattapone, S. *In vitro* amplification of protease-resistant prion protein requires free sulphydryl groups. *Biochemistry* **42**, 4127–4135 (2003).
4. Saborio, G. P., Permanne, B. & Soto, C. Sensitive detection of pathological prion protein by cyclic amplification of protein misfolding. *Nature* **411**, 810–813 (2001).
5. Lockard, R. E. & Kumar, A. Mapping tRNA structure in solution using double-strand-specific ribonuclease V1 from cobra venom. *Nucleic Acids Res.* **9**, 5125–5140 (1981).
6. Banks, G. R. A ribonuclease H from *Ustilago maydis*. Properties, mode of action and substrate specificity of the enzyme. *Eur. J. Biochem.* **47**, 499–507 (1974).
7. Kocisko, D. A. *et al.* Cell-free formation of protease-resistant prion protein. *Nature* **370**, 471–474 (1994).
8. Caughey, B., Horiuchi, M., Demaimay, R. & Raymond, G. J. Assays of protease-resistant prion protein and its formation. *Methods Enzymol.* **309**, 122–133 (1999).
9. Derrington, E. *et al.* PrPC has nucleic acid chaperone properties similar to the nucleocapsid protein of HIV-1. *C. R. Acad. Sci. III* **325**, 17–23 (2002).
10. Moscardini, M. *et al.* Functional interactions of nucleocapsid protein of feline immunodeficiency virus and cellular prion protein with the viral RNA. *J. Mol. Biol.* **318**, 149–159 (2002).
11. Gabus, C. *et al.* The prion protein has RNA binding and chaperone properties characteristic of nucleocapsid protein NCP7 of HIV-1. *J. Biol. Chem.* **276**, 19301–19309 (2001).
12. Gabus, C. *et al.* The prion protein has DNA strand transfer properties similar to retroviral nucleocapsid protein. *J. Mol. Biol.* **307**, 1011–1021 (2001).
13. Nandi, P. K., Leclerc, E., Nicole, J. C. & Takahashi, M. DNA-induced partial unfolding of prion protein leads to its polymerisation to amyloid. *J. Mol. Biol.* **322**, 153–161 (2002).
14. Cordeiro, Y. *et al.* DNA converts cellular prion protein into the β -sheet conformation and inhibits prion peptide aggregation. *J. Biol. Chem.* **276**, 49400–49409 (2001).
15. Weissmann, C. A 'unified theory' of prion propagation. *Nature* **352**, 679–683 (1991).
16. Chapon, C., Cech, T. R. & Zaug, A. J. Polyadenylation of telomerase RNA in budding yeast. *RNA* **3**, 1337–1351 (1997).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements The authors thank G. Saborio, C. Soto, V. Ambros, C. Cole and W. Wickner for helpful advice. This work was supported by the Burroughs Wellcome Fund Career Development Award, the Hitchcock Foundation, and an NIH Clinical Investigator Development Award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.S. (supattapone@dartmouth.edu).

Loss of Omi mitochondrial protease activity causes the neuromuscular disorder of *mnd2* mutant mice

Julie M. Jones¹, Pinaki Datta⁴, Srinivasa M. Srinivasula⁴, Weizhen Ji¹, Sanjeev Gupta⁴, ZhiJia Zhang⁴, Erika Davies⁵, György Hajnóczky⁵, Thomas L. Saunders², Margaret L. Van Keuren³, Teresa Fernandes-Alnemri⁴, Miriam H. Meisler¹ & Emad S. Alnemri⁴

¹Department of Human Genetics, ²Department of Internal Medicine, Division of Molecular Medicine and Genetics, and ³Transgenic Animal Model Core, Biomedical Research Core Facilities, University of Michigan, Ann Arbor, Michigan 48109-0618, USA

⁴Center for Apoptosis Research and the Department of Microbiology and Immunology, Kimmel Cancer Institute, and

⁵Department of Pathology and Cell Biology, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

The mouse mutant *mnd2* (motor neuron degeneration 2) exhibits muscle wasting, neurodegeneration, involution of the spleen and thymus, and death by 40 days of age^{1,2}. Degeneration of striatal neurons, with astrogliosis and microglia activation, begins at around 3 weeks of age, and other neurons are affected at later stages³. Here we have identified the *mnd2* mutation as the missense mutation Ser276Cys in the protease domain of the nuclear-encoded mitochondrial serine protease Omi (also

known as HtrA2 or Prss25). Protease activity of Omi is greatly reduced in tissues of *mnd2* mice but is restored in mice rescued by a bacterial artificial chromosome transgene containing the wild-type Omi gene. Deletion of the PDZ domain partially restores protease activity to the inactive recombinant Omi protein carrying the Ser276Cys mutation, suggesting that the mutation impairs substrate access or binding to the active site pocket. Loss of Omi protease activity increases the susceptibility of mitochondria to induction of the permeability transition, and increases the sensitivity of mouse embryonic fibroblasts to stress-induced cell death. The neurodegeneration and juvenile lethality in *mnd2* mice result from this defect in mitochondrial Omi protease.

mnd2 was identified in 1990 as a spontaneous, recessively inherited mutation that arose on the C57BL/6J inbred background¹. The earliest symptoms are altered gait and cessation of normal weight gain, followed by ataxia, repetitive movements and akinesia. Striatal neurons are most susceptible, but other neurons in the central nervous system, brain stem and spinal cord, including motor neurons, are affected at later stages^{1,3}. The neuronal degeneration combines features of necrosis with features of apoptosis including nuclear condensation, TdT-mediated dUTP nick end labelling (TUNEL) staining, and DNA laddering³. Early degeneration of mitochondria was also noted³.

To identify the molecular basis for the *mnd2* disorder, we crossed the mutant strain C57BL/6J-*mnd2* with CAST/Ei². Analysis of 1,208 F₂ mice localized the gene to a 0.9-megabase (Mb) interval of mouse chromosome 6 (Fig. 1a). The human orthologue is located in a

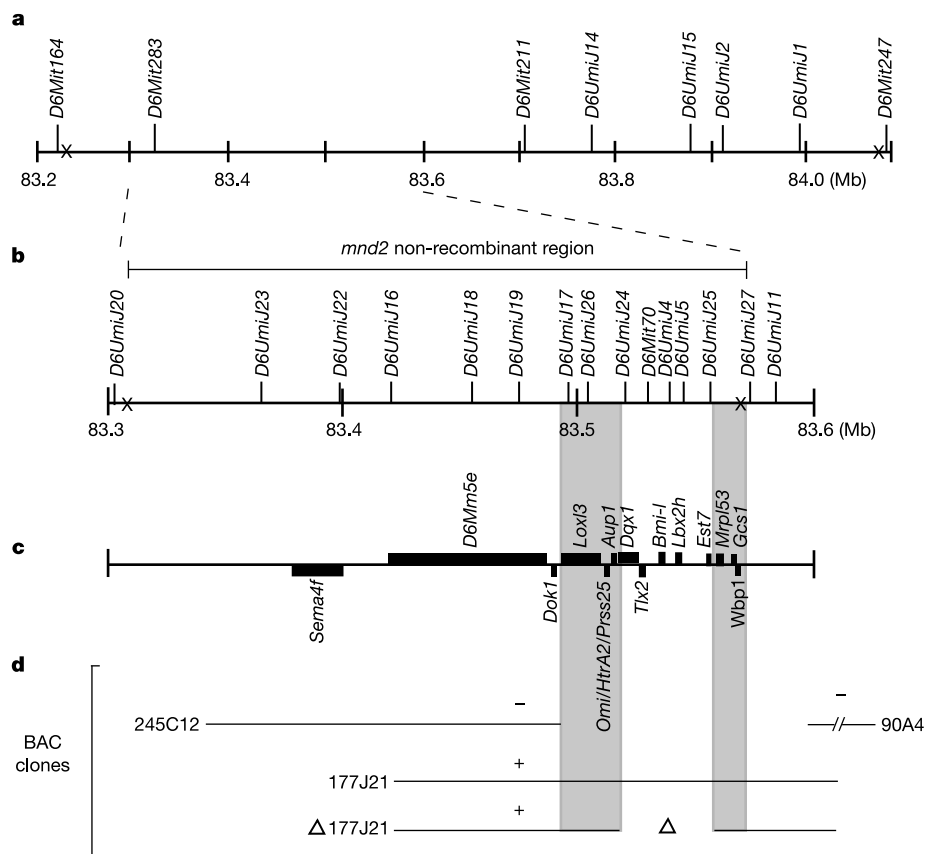


Figure 1 Genetic and physical map of the *mnd2* region. **a**, Previous non-recombinant region². **b**, Reduced non-recombinant interval from the cross between C57BL/6J-*mnd2* and SJL/J mice. **c**, Gene content of the non-recombinant region. Genes above and below the horizontal line have opposite orientation on the chromosome. **d**, Overlapping BAC

clones of mouse genomic DNA spanning the non-recombinant region. A plus sign indicates BACs that rescue the *mnd2* phenotype in transgenic mice; a minus sign indicates BACs that do not rescue the *mnd2* phenotype; a triangle indicates the position of the BAC 177J21 deletion in transgenic line Tg410.

Table 1 Transgenic rescue of *mnd2* homozygotes by expression of BAC clones

BAC clone	Tg line	Number of rescued mice (<i>mnd2/mnd2</i> ;Tg/+)
177J21	Tg514	21 of 21
Δ177J21	Tg410	10 of 10
90A4	Tg70	0 of 5
245C12	Tg666, 673	0 of 5

Clones are described in Fig. 1d. Each transgenic line is derived from an individual founder carrying the indicated BAC clone. Non-transgenic *mnd2/mnd2* mice are severely impaired and do not survive beyond 40 days of age. The oldest rescued mice have currently survived beyond 10 months with no phenotypic abnormalities.

gene-dense region of human chromosome 2p13 with conserved gene order and content⁴. Because of the low rate of recombination in the original cross, we initiated a second cross between strains C57BL/6J-*mnd2* and SJL/J. Recombination breakpoints in 400 F₂ mice from the second cross localized *mnd2* to an interval of 250 kilobases (kb) containing 14 genes (Fig. 1b, c).

Three BAC clones containing mouse genomic DNA spanning the non-recombinant region were tested for their ability to rescue the lethal *mnd2* phenotype (Fig. 1d). Transgenic mice carrying each BAC clone were generated by microinjection of fertilized eggs. The *mnd2* phenotype was completely corrected in transgenic line Tg514 carrying BAC clone 177J21 (Table 1). Homozygous *mnd2/mnd2*

mice carrying this transgene have normal weight gain and fertility, and do not exhibit any of the phenotypic abnormalities of the *mnd2* disorder. BAC clones 245C12 and 90A4 did not rescue *mnd2* homozygotes. In transgenic line Tg410, a deletion occurred within BAC 177J21, but the ability to rescue *mnd2* mice was retained (Table 1). The size and position of the deletion, Δ177J21, was determined by crossing the transgene onto strain SJL and testing for C57BL/6J alleles of polymorphic *D6UmiJ* markers (Fig. 1d). Combining information from the deleted BAC clone and the distal recombination breakpoint localized *mnd2* to a 40-kb region containing six candidate genes: *Lox13*, *Omi*, *Aup1*, *Mrpl53*, *Gcs1* and *Wbp1* (Fig. 1, shaded areas).

In parallel with the mapping and rescue experiments, we compared the mutant and wild-type sequences for these six candidate genes as well as most of the other genes in the original 0.9-Mb region⁴⁻⁸. Comparison of >50 kb of gene sequences from *mnd2* and C57BL/6J identified only one difference, an A to T nucleotide substitution in exon 3 of the nuclear-encoded mitochondrial serine protease Omi (Fig. 2a). This missense mutation changes serine 276 to cysteine. Serine 276 is located in the protease domain of Omi close to serine 306, the active site serine⁹, and is evolutionarily conserved in prokaryotic and eukaryotic homologues (Fig. 2b).

Omi is localized in the intermembrane space of the mitochondria¹⁰⁻¹⁴ and is a member of the HtrA protease-chaperone family.

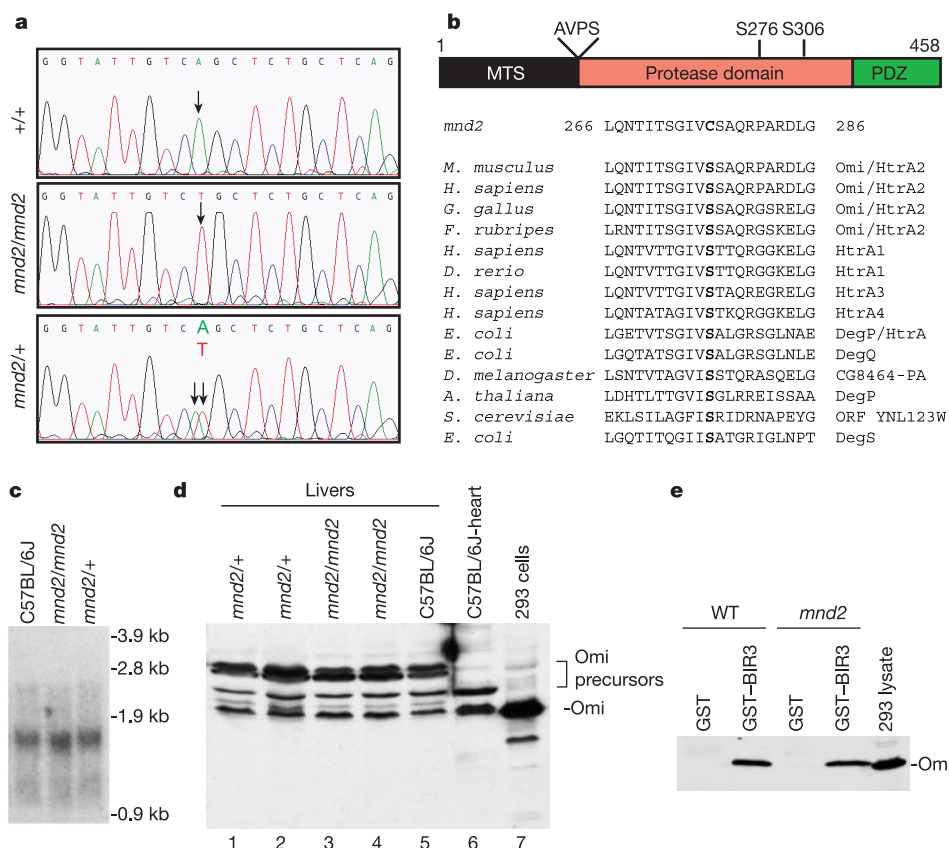


Figure 2 Ser276Cys mutation in Omi. **a**, Nucleotide sequence chromatogram from exon 3 of wild-type (top panel), *mnd2* (middle panel) and heterozygous (bottom panel) DNA. The *Prss25^{mnd2}* mutation, an A to T transversion at nucleotide 826, is indicated by the arrow. **b**, Evolutionary conservation of Ser 276 in HtrA proteins. The AVPS motif at residues 134–137 binds to IAPs. Ser 276 and the active site Ser 306 in the protease domain are indicated. MTS, mitochondrial localization signal. **c**, Northern blot with 5 μg of poly(A)⁺ RNA was hybridized with the 0.8-kb *EcoRI/NotI* fragment from expressed sequence tag clone 1264040 (GenBank AA881220). **d**, Western blot of Omi protein in total liver extracts

from five mice of indicated genotypes. Lanes 6 and 7 contain total extracts from wild-type mouse heart and 293T human cells, respectively. All lanes contain equal amounts of total proteins. **e**, Western blot analysis of glutathione *S*-transferase (GST)–BIR3-bound Omi proteins from wild-type and *mnd2* liver extracts. Liver extracts from wild-type (WT) or *mnd2* mice were bound to GST (GST lanes) or GST–BIR3 (GST–BIR3 lanes) fusion protein immobilized on glutathione-Sepharose beads. The beads were washed and the bound proteins were eluted with SDS sample buffer and western-blotted with anti-Omi-specific antibody. Last lane, human 293T cells.

Bacterial members are quality control proteases that are essential for surviving high temperature or oxidative stress^{15,16}. Recently, Omi was shown to be released from the mitochondria with cytochrome *c* and Smac (also known as DIABLO) during apoptosis^{10–14}. The released Omi can bind to inhibitors of apoptosis proteins (IAPs), inhibit their activity and directly degrade them^{17,18}.

The missense mutation does not alter the size or steady-state abundance of the Omi messenger RNA in *mnd2* tissues (Fig. 2c). Evaluation of the Omi protein also revealed no effect on the molecular mass or protein concentration (Fig. 2d). The precursor protein appears to be correctly processed into the mature IAP-binding form by protease cleavage of the amino terminus in mutant tissues (Fig. 2d). Binding of Omi to the Baculovirus Inhibitory Repeat 3 (BIR3) domain of X-chromosome-linked IAP (XIAP) was comparable in wild-type and mutant tissue extracts (Fig. 2e), indicating that maturation and the IAP-binding activity of Omi are not affected by the Ser276Cys substitution. To determine the effect on protease activity, Omi was immunoprecipitated from liver

and assayed with ³⁵S-labelled β-casein as substrate (Fig. 3a). No β-casein cleaving activity was detectable in the immunoprecipitates from *mnd2* tissues (Fig. 3a). In livers of *mnd2/mnd2* homozygotes from lines Tg410 and Tg514, which are phenotypically rescued by BAC clones 177J21 and Δ177J21, the protease activity of Omi is restored (Fig. 3b).

The physiological concentration of Omi in highly expressing cells such as 293T cells is approximately 150 nM (Fig. 2d, lane 7; see also ref. 17). Recombinant Omi containing the Ser276Cys mutation had no detectable protease activity at concentrations of up to seven times this concentration (Fig. 3c, lane 5). A low level of residual activity was only detectable at much higher concentrations (Fig. 3c, lane 6). The mutant protein was also inactive towards the substrates XIAP and c-IAP1 (Fig. 3d).

Although the Ser276Cys mutant does not have protease activity at physiological concentrations, the results in Figs 2 and 3 indicate that it is correctly processed into the mature form. These observations suggest that processing of Omi into the mature form is probably

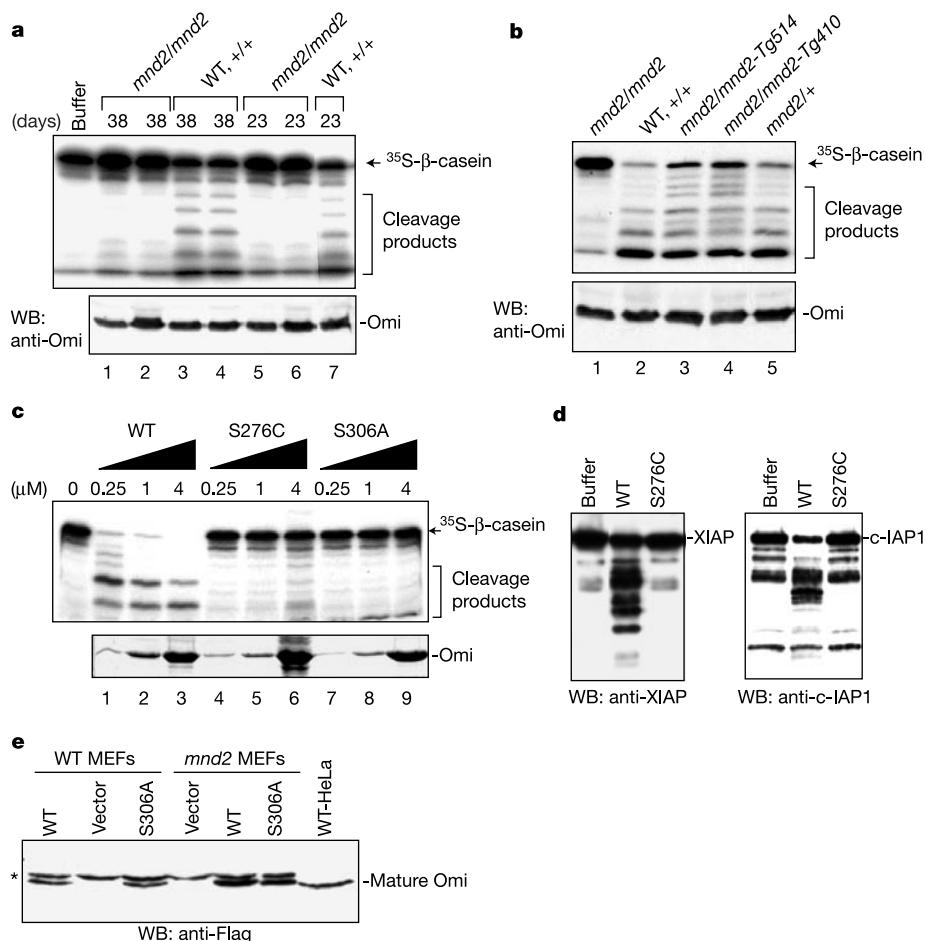


Figure 3 Effect of the Ser276Cys mutation on the protease activity of Omi. **a**, Loss of serine protease activity of Omi in *mnd2* tissues. Top panel: endogenous Omi was immunoprecipitated from liver extracts of 23- and 38-day-old mice and assayed with ³⁵S-labelled β-casein as substrate. Bottom panel: western blot of liver S100 extracts with Omi-specific polyclonal antibody demonstrating normal amounts of protein in *mnd2*. **b**, The Omi activity is restored in transgenic lines Tg514 and Tg410. **c**, Loss of activity of recombinant Omi with the Ser276Cys mutation at near physiological concentration. Top panel: ³⁵S-labelled β-casein was incubated with increasing concentrations of affinity-purified recombinant Omi. Cleavage products were analysed by SDS-PAGE and autoradiography. Bottom panel: Coomassie-stained gel of affinity-purified Omi proteins.

The active site mutant Ser306Ala lacking activity is included as control. **d**, Western blot analysis of IAP cleavage by recombinant Omi protein. Purified XIAP (left panel) or c-IAP1 (right panel) were incubated with buffer or 250 nM Omi protein. Products were separated by SDS-PAGE and immunoblotted with antibodies against XIAP or c-IAP1 as indicated. **e**, Maturation of Omi is not affected by the Ser276Cys mutation. MEFs were transfected with constructs encoding C-terminal Flag-tagged Omi precursors. At 24 h after transfection, cells were solubilized in SDS sample buffer, fractionated by SDS-PAGE, and immunoblotted with anti-Flag antibody. Last lane, lysate from HeLa cells expressing mature C-terminal Flag-tagged wild-type Omi. Asterisk, nonspecific band. The lanes labelled buffer contain substrate incubated with buffer alone.

catalysed by other mitochondrial proteases. To test this hypothesis, we overexpressed wild-type Omi and the active site mutant Ser306Ala in mouse embryonic fibroblast cells (MEFs). If Omi is involved in its own processing, then MEFs from *mnd2* mice should be less effective in processing the inactive Ser306Ala protein. No difference was observed in the extent of processing in wild-type and *mnd2* MEFs (Fig. 3e), indicating that other mitochondrial proteases are involved in maturation of Omi.

In the recently published X-ray structure of Omi, Ser 276 is located near a region critical for homotrimerization⁹. As formation of a homotrimeric complex is required for protease activity, we evaluated the effect of the mutation on trimerization. Flag- and T7-tagged Omi proteins were co-expressed in 293T cells and cell extracts were prepared. T7-tagged and endogenous Omi proteins were co-immunoprecipitated with the Flag-tagged Ser276Cys protein (Fig. 4a), indicating that the mutation does not prevent homotrimerization.

The PDZ domain in Omi restricts access of substrate to the active site pocket⁹. The location of Ser 276 in the highly flexible L3 loop, which is flanked by the PDZ domain on one side and the active site pocket on the opposite side^{9,15}, suggested that the Ser276Cys mutation might negatively affect the packing interactions at the interface between PDZ and protease domains, resulting in loss of access to the active site pocket. Deletion of the PDZ domain increased the activity of the wild-type protein and partially restored the protease activity of the Ser276Cys mutant (Fig. 4b). The Ser276Cys mutation may affect flexibility of the L3 loop, forcing the PDZ domain into a conformation that blocks the active site pocket of the mutant protein. Alternatively, the Ser276Cys substitution may destabilize intramolecular interactions between the PDZ and protease domains, preventing the PDZ domain from orienting the substrate into the active site pocket. Substrate binding to the

active site may also be impaired, as deletion of the PDZ domain did not fully restore protease activity.

The *mnd2* MEFs lack Omi activity (Fig. 5a). To examine the effect of loss of Omi protease activity on cell viability, *mnd2* and wild-type MEFs were treated with stress-inducing agents and subsequently stained with annexin V and propidium iodide. The proportion of stained cells was greatest in homozygous *mnd2/mnd2* MEFs (Fig. 5b), indicating increased cell death. The *mnd2* MEFs exhibited annexin V and propidium iodide staining (Fig. 5c), suggesting that both apoptosis (annexin V) and necrosis (propidium iodide) occur in these cells. This is consistent with observations of striatal neurons in *mnd2* mice, which combine aspects of apoptosis and necrosis³. The caspase inhibitor z-VAD-fmk did not completely inhibit death of *mnd2* MEFs (Fig. 5d), suggesting that some cell death occurs by a caspase-independent mechanism. The data indicate that loss of Omi protease activity increases sensitivity to stress-induced cell death.

Induction of the mitochondrial permeability transition can cause mitochondrial membrane permeabilization and dysfunction leading to apoptosis and necrosis¹⁹. To determine whether the Ser276Cys substitution in Omi affects the integrity of the mitochondrial membrane barrier, fluorimetric measurement of the mitochondrial membrane potential ($\Delta\Psi_m$) and cytosolic Ca^{2+} concentration ($[Ca^{2+}]_c$) was carried out in permeabilized MEFs after calcium pulsing. In surviving cells, Ca^{2+} is taken up by mitochondria during $[Ca^{2+}]_c$ spiking and is used to stimulate mitochondrial ATP production. However, in cells exposed to certain pro-apoptotic agents, propagation of $[Ca^{2+}]_c$ signals to the mitochondria triggers activation of the permeability transition pore (PTP) and mitochondrial membrane permeabilization, leading to execution of apoptosis²⁰. In wild-type MEFs, added Ca^{2+} caused only a small and transient depolarization and was effectively removed from cytosol by the mitochondria (Fig. 6, black traces and bars). In contrast, in *mnd2* MEFs, Ca^{2+} pulsing induced relatively large and prolonged mitochondrial depolarization and a progressive rise in $[Ca^{2+}]_c$ (Fig. 6, grey traces and bars). These results suggest that *mnd2* mutant cells are more vulnerable to Ca^{2+} -induced permeability transition and mitochondrial membrane permeabilization. The Ser276Cys substitution in Omi thus seems to promote activation of the PTP by Ca^{2+} . As stress-inducing agents may also target the PTP through elevation in Ca^{2+} concentration or free radical formation, sensitization of the PTP in *mnd2* mutant cells provides a potential mechanism to explain the increased sensitivity of *mnd2* MEFs to stress-induced cell death, as well as the degeneration of striatal neurons in *mnd2* mice. Supporting this mechanism, recent studies have shown that striatal mitochondria are more vulnerable to calcium-induced permeability transition than cortical mitochondria¹⁹. Therefore, mitochondrial dysfunction is probably responsible for the selective death of striatal neurons in *mnd2* mice.

We have demonstrated that the Ser276Cys missense mutation in Omi is responsible for neurodegeneration in *mnd2* mutant mice. Omi is a serine protease normally present in the intermembrane space of mitochondria^{10–14}. *In vitro* studies with cancer cell lines have shown that release of Omi from the mitochondria into the cytosol causes inhibition and degradation of IAPs and apoptosis^{17,18}. However, there are no loss-of-function models that support a role for Omi in apoptosis *in vivo*. On the contrary, the data presented here demonstrate that loss of the protease activity of Omi increases sensitivity to stress-induced cell death and is probably responsible for the massive loss of striatal neurons in *mnd2* mutant mice³. The data can be explained if Omi has two distinct functions: (1) under normal physiological conditions it is required within mitochondria for maintenance of mitochondrial homeostasis, and (2) under apoptotic conditions it is released from mitochondria and has an apoptotic role that is shared by other released proteins such as cytochrome c (ref. 21) and apoptosis-inducing factor (AIF)^{22,23}. We

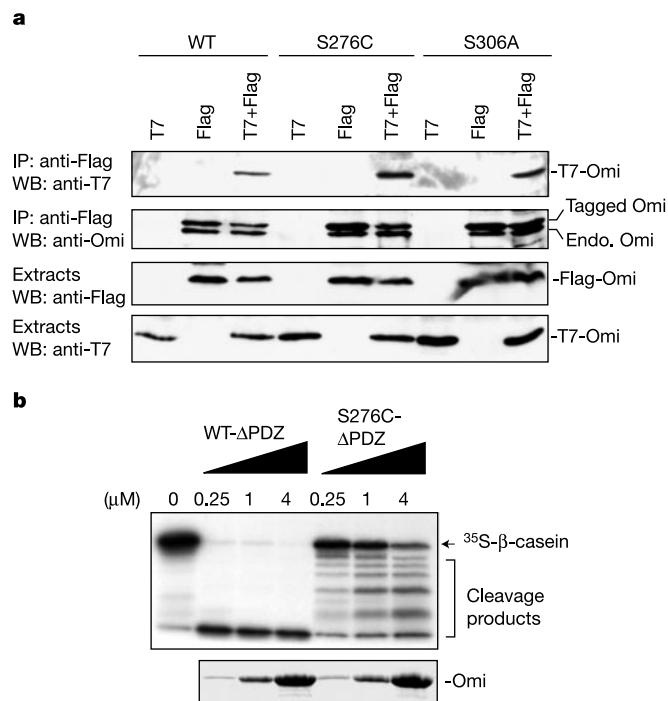


Figure 4 Structural determinants of Omi activity. **a**, The Ser276Cys mutation does not affect homotrimerization. (For assay conditions, see Methods section.) **b**, Deletion of the PDZ domain partially restores the protease activity of the Ser276Cys Omi mutant. The activity of the PDZ-deleted Omi proteins was determined as in Fig. 3c. Wild-type PDZ-deleted protein catalysed the complete digestion of substrate.

propose that the Ser276Cys mutation compromises the mitochondrial function, leading to mitochondrial membrane permeabilization and cytoplasmic release of mutant Omi along with other active inducers of cell death.

The structural similarity between Omi and bacterial HtrA proteins^{15,16,24} suggests that Omi may be a sensor of unfolding stresses in the mitochondria. Thus, loss of Omi protease activity in the *mnd2* mice could lead to accumulation of misfolded and damaged proteins in the mitochondria, increased susceptibility to induction of the permeability transition and mitochondrial dysfunction. Progressive mitochondrial dysfunction can lead to irreversible loss of neuronal cells, and this is thought to be an important pathogenic mechanism in many neurodegenerative disorders. Mitochondrial disintegration is an early component of neuronal degeneration in the striatum of *mnd2* mice³. The dying striatal neurons exhibit features of apoptosis such as condensation of nuclear DNA, nuclear shrinkage, TUNEL staining and DNA laddering³. Some cells also exhibit features of necrosis³. This neuronal degradation was not prevented by overexpression of Bcl2 (ref. 3). The loss of protein quality control in the mitochondria can account for both types of cell death in *mnd2* mice. Induction of mitochondrial permeability

transition due to accumulation of misfolded proteins might lead to release of the mitochondrial apoptotic proteins that activate the caspase-dependent and -independent cell death cascades. At the same time, the loss of mitochondrial respiratory competence, and reduced intracellular ATP levels, might result in rupture of the plasma membrane and necrotic cell death²⁵.

AIF is a nuclear-encoded mitochondrial protein that is localized in the intermembrane space and is involved in apoptosis²⁶. In the mouse mutant harlequin (*Hq*), AIF expression is reduced to 20% of normal, resulting in oxidative stress and degeneration of cerebellar neurons²³. The mechanism of cell death in this mutant is loss of the hydrogen peroxide scavenger function of AIF. BAD is another protein involved in apoptosis, which has recently been shown to reside in a mitochondrial complex that integrates glycolysis and apoptosis²⁷. Similar to Omi, these proteins combine an essential mitochondrial function with a role in inducing apoptosis. The human neurodegenerative disease hereditary spastic paraplegia (SPG7) is also caused by inactivation of a mitochondrial protease called paraplegin²⁸. The yeast homologues of paraplegin, known as AAA proteases, are quality control protease-chaperons anchored to the inner mitochondrial membrane²⁹. In both SPG7 and *mnd2* the

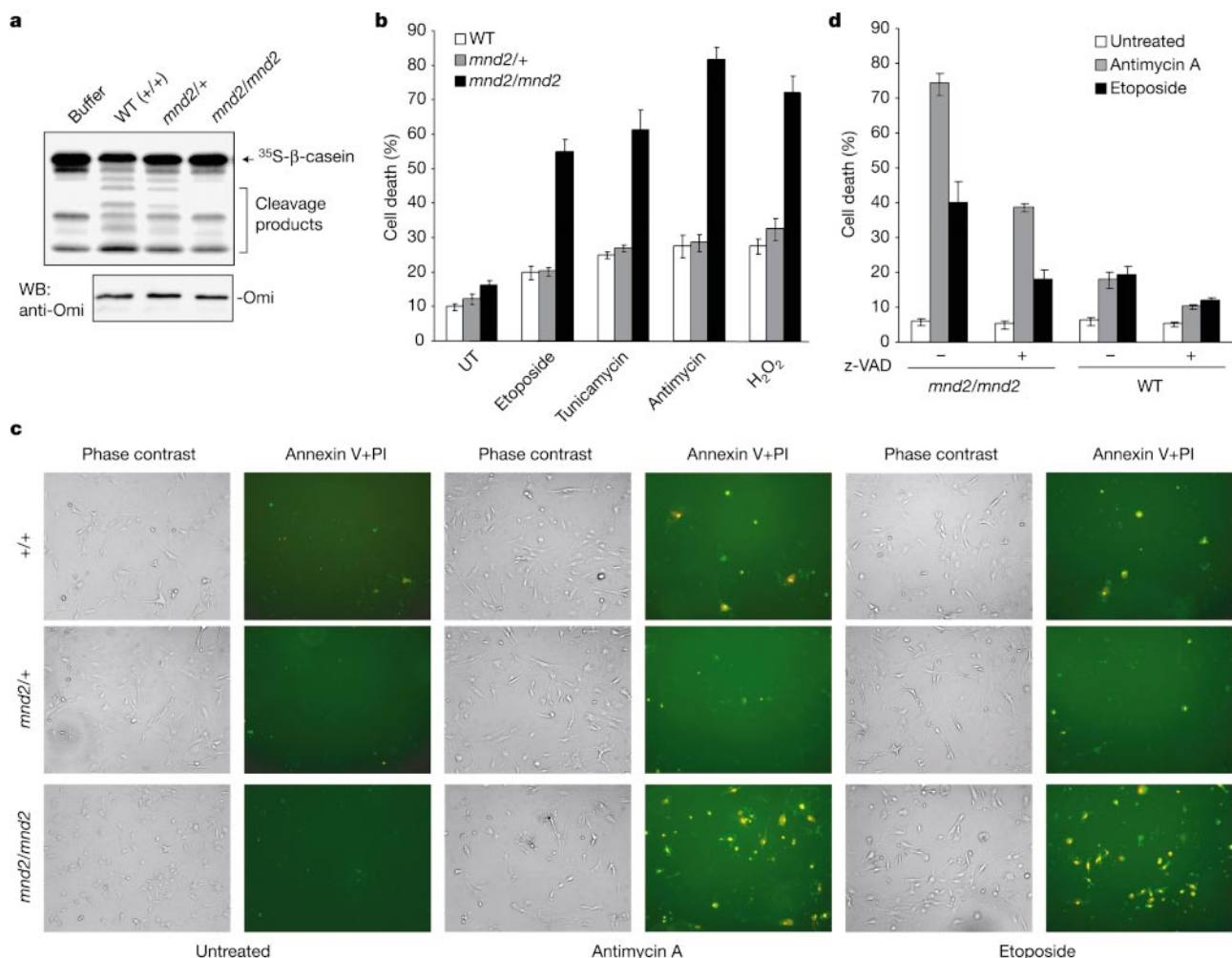


Figure 5 Sensitivity of *mnd2* MEFs to stress-induced cell death. **a**, The top panel shows protease activity of Omi in cultured MEFs. Activity was determined as in Fig. 3a. The bottom panel shows western blot of MEF extracts with Omi-specific antibody. **b**, MEFs were treated with stress-inducing agents and then stained with annexin V and propidium iodide. The percentage of cell death was determined as described in Methods. UT, untreated. **c**, Representative fields of control (untreated) MEFs and treated MEFs

stained with annexin V (green fluorescence) and propidium iodide (PI, orange fluorescence). Cells were visualized by fluorescence microscopy ($\times 10$). **d**, z-VAD-fmk partially inhibits stress-induced cell death of *mnd2* MEFs. MEFs were treated with antimycin A or etoposide in the presence (+) or absence (-) of the pan-caspase inhibitor z-VAD-fmk.

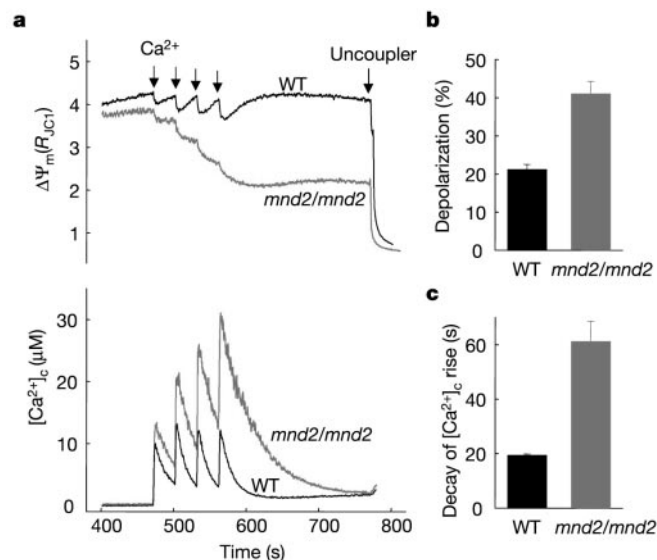


Figure 6 Ca^{2+} pulsing-induced mitochondrial membrane permeabilization in *mnd2* MEFs. **a**, $\Delta\Psi_m$ (top) and $[\text{Ca}^{2+}]_i$ (bottom) recordings from *mnd2/mnd2* (grey) and wild-type (black) MEFs. Additions: Ca^{2+} , 30 μM CaCl_2 at each arrow; uncoupler, 1 μM FCCP. Data are representative of 5–6 different measurements. **b**, Quantification of mitochondrial depolarization in response to Ca^{2+} pulsing. To quantify Ca^{2+} -induced depolarization, the Ca^{2+} -induced decrease in the JC1 fluorescence ratio R_{JC1} was normalized to the uncoupler-induced decrease in R_{JC1} in both *mnd2/mnd2* and wild-type MEFs. **c**, Determination of mitochondrial Ca^{2+} uptake. To evaluate mitochondrial Ca^{2+} uptake, the duration at half maximum of the $[\text{Ca}^{2+}]_i$ transient evoked by the fourth Ca^{2+} pulse was calculated. Data from separate experiments are shown as mean $\pm$ s.e.m.

cause of mitochondrial dysfunction and neurodegeneration may include defective protein quality control, in the mitochondrial matrix or intermembrane space, respectively. It remains to be seen whether spontaneous mutations in Omi are responsible for neurodegeneration disorders in humans.

Note added in proof: The human *OMI* gene is located near the Parkinson's disease susceptibility locus *PARK3* on chromosome 2p13.1. No coding or splice site mutations of human *OMI* were detected in two families with Parkinson's disease mapped to the *PARK3* locus (J.M.J., S.M.S., T.F.-A., M.H.M. and E.S.A., unpublished observations). □

Methods

Animals

The *mnd2* mutation arose on strain C57BL/6J in 1990 at the University of Michigan. C57BL/6J-*mnd2*/+ males were mated with SJL/J females. The F_1 offspring were test-crossed with C57BL/6J-*mnd2*/+ mice. Confirmed heterozygotes were intercrossed to generate F_2 mice. C57BL/6J and SJL/J mice were obtained from the Jackson Laboratory.

BAC transgenic mice

BAC clones RP23-177J21 and RP23-90A4 were obtained from BAC/PAC resources (Oakland Children's Hospital); clone 245C12 has been described previously². DNA from BAC clones was purified with Nucleobond columns (BD Biosciences-Clontech) (<http://www.med.umich.edu/tamc/BACnucleo.html>). Female F_2 heterozygotes from the SJL cross were superovulated and mated with F_1 heterozygous males to generate fertilized eggs for microinjection. Transgenic founders were identified by polymerase chain reaction (PCR) amplification of tail DNA using BAC-specific primers flanking each vector-insert junction. The integrity of the BAC transgenes 177J21 and 90A4 was analysed by crossing onto a homozygous SJL/J background and detection of C57BL/6J alleles of the appropriate *D6Umi* markers (Fig. 1a, b). The integrity of BAC 245C12 was evaluated by detection of alleles from strain 129/Sv on a C57BL/6J background.

Markers and genotyping

Nineteen new genetic markers, *D6Umi*1 to *D6Umi*27, with different alleles in C57BL/6J and SJL/J were generated using primers flanking microsatellite sequences in the C57BL/6J genome. Primer sequences and allele sizes are available from the Mouse Genome Database (www.informatics.jax.org). The Ser276Cys mutant allele of Omi, designated *Prss25^{mnd2}*, is

detected by amplification of a 500-base pair (bp) fragment containing exon 3 from genomic DNA using the primers *mnd2F* (5'-CACACTGAGGATTCAAACCAAGGT-3') and *mnd2R* (5'-GACCGAGGACATAAACAGGGTGTGTA-3'). Digestion with *AluI* produces a 244-bp product from the mutant allele in place of wild-type fragments of 171 bp and 73 bp. PCR of genomic DNA was carried out using the primers described in the text, in a volume of 25 μl containing $\times 1$ Taq DNA polymerase buffer (Promega), 2.5 mM MgCl_2 , 0.2 mM dNTPs, 0.5 μM primers and 1 U Taq DNA polymerase in storage buffer B (Promega). Incubation at 94 °C for 2 min was followed by 34 cycles of 94 °C for 45 s, 60 °C for 45 s, and 72 °C for 45 s, followed by 72 °C for 10 min. PCR products were separated on 2% agarose gels and stained with ethidium bromide.

DNA sequence

The DNA sequences of BAC clones RP23-177J21 and b245C12 are described in GenBank under accession numbers AC104324 and AC003061, respectively. Exons of positional candidate genes were amplified from wild-type and *mnd2/mnd2* samples by PCR or PCR with reverse transcription (RT) and sequenced by the University of Michigan DNA Sequencing Core.

Complementary DNA cloning and recombinant protein expression

The wild-type full-length mouse Omi cDNA clone (GenBank AF175324) was obtained from A. Zervos. Constructs encoding wild-type and mutant full-length and mature Omi proteins were generated by PCR and overlapping PCR, using modified complementary PCR adaptor-primers. Carboxy-terminal Flag, T7 or (His)6 tagging was carried out by cloning the PCR-generated cDNAs in-frame into Flag-C-pcDNA3, T7-C-pcDNA3 or pET-21 expression vectors respectively. The human β -casein cDNA was isolated from human placenta by RT-PCR and cloned into pcDNA3 to generate the pcDNA- β -casein plasmid. All expression constructs were sequenced by the Kimmel Cancer Institute DNA sequencing core. Mature mouse Omi was overexpressed in *Escherichia coli* strain BL21(DE3) as a C-terminally (His)6-tagged protein using a pET-21 vector. ³⁵S-labelled β -casein protein was prepared by *in vitro* translation in the presence of ³⁵S-methionine using the TNT rabbit reticulocyte lysate kit from Promega. XIAP and c-IAP1 were expressed in bacteria as described^{10,17}.

GST-BIR3 pull-down assay

This assay was performed essentially as described¹⁰.

Protease assay

The protease activity of recombinant wild-type and mutant Omi was assayed with the substrates ³⁵S-labelled β -casein, c-IAP1 or XIAP as described previously¹⁷. Different amounts of bacterially expressed (His)6-tagged Omi were purified on Talon affinity resins and then incubated with ³⁵S-labelled β -casein or IAPs in an assay buffer containing 25 mM Tris, pH 7.6, 100 mM NaCl, 250 mM imidazole and 1 mM dithiothreitol for 1 h at 37 °C in a total 50- μl reaction. Cleavage products were analysed by SDS-polyacrylamide gel electrophoresis (PAGE) and visualized by autoradiography or western blotting.

Endogenous Omi proteins were isolated by immunoprecipitation before protease assay. Frozen livers from *mnd2* and control mice were homogenized in a buffer containing 50 mM Tris, pH 7.6, 150 mM NaCl, 0.1% NP-40 and protease inhibitors, and centrifuged at 20,000g. The supernatants were further centrifuged at 100,000g and the resulting S100 extracts were incubated with an Omi-specific polyclonal antibody¹⁷ for 1 h. After incubation, the Omi-antibody complexes were bound to protein G Sepharose and washed several times. The protein G Sepharose-bound complexes were incubated with ³⁵S-labelled β -casein in an assay buffer (25 mM HEPES, pH 7.6, 10 mM KCl, 1.5 mM MgCl_2 , 1 mM EDTA, 1 mM EGTA) for 1 h at 37 °C in a total 50 μl reaction. Cleavage products were analysed by SDS-PAGE and visualized by autoradiography. Omi activity in MEFs was assayed by the same method.

Trimerization assay

293T cells were co-transfected with constructs encoding C-terminal Flag-tagged and T7-tagged Omi variants (wild type, S276C, S306A) using the Lipofectamine method. At 36 h after transfection, cells were lysed and the cellular lysates were immunoprecipitated with a Flag-specific antibody. The immunoprecipitates were analysed by western blotting with a T7-specific or Omi-specific antibodies.

Cell death assay

MEFs were grown in DMEM supplemented with 10% FBS and penicillin/streptomycin. MEFs were seeded in 12-well plates at a density of 5×10^4 cells per well and treated with antimycin A (100 $\mu\text{g ml}^{-1}$ for 12 h), hydrogen peroxide (3 μM for 5 h), etoposide (100 $\mu\text{g ml}^{-1}$ for 12 h), or tunicamycin (2.5 $\mu\text{g ml}^{-1}$ for 12 h) in the presence or absence of z-VAD-fmk (50 μM). Cells were stained with annexin V and propidium iodide (CLONTECH Laboratories) as per the manufacturer's recommendations and counted. Normal cells and dead cells (annexin V and/or propidium iodide positive) were counted using fluorescence microscopy. The percentage of dead cells in each experiment was expressed as the mean percentage of dead cells as a fraction of the total number of cells counted. Data represent mean values from three experiments.

$\Delta\Psi_m$ and $[\text{Ca}^{2+}]_i$ assay

Measurement of $\Delta\Psi_m$ and $[\text{Ca}^{2+}]_i$ in permeabilized MEFs was carried out as described previously²⁰. Equal aliquots of cells (2.4 mg protein) were resuspended and permeabilized with 40 $\mu\text{g ml}^{-1}$ digitonin in 1.5 ml of an intracellular medium composed of 120 mM KCl, 10 mM NaCl, 1 mM KH_2PO_4 , 20 mM HEPES/Tris, pH 7.2 supplemented with 1 $\mu\text{g ml}^{-1}$ of each of the following: antipain, leupeptin, pepstatin. Measurements were carried out in

the presence of 2 mM MgATP, 5 mM phosphocreatine, 5 U ml⁻¹ creatine kinase and 2 mM succinate.

To carry out simultaneous measurements of [Ca²⁺]_i and ΔΨ_m, the permeabilized cells were supplemented with 0.5 μM fura2FF/FA and 800 nM JC1. Fluorescence was monitored in a multi-wavelength-excitation dual wavelength-emission fluorimeter (DeltaRAM, PTI) using 340- and 380-nm excitation and 535-nm emission for fura2FF whereas 490-nm excitation/535-nm emission and 570-nm excitation/595-nm emission were used for JC1. ΔΨ_m is shown as the ratio of the fluorescence of J-aggregate (excitation 570 nm/emission 595 nm) and monomer (excitation 490 nm/emission 535 nm) forms of JC1. Calibration of fura2FF signal was carried out by adding CaCl₂ (2.5 mM) and subsequently EGTA/Tris (10 mM) at pH 8.5. Experiments were carried out at 35 °C and with simultaneous stirring.

Received 31 July; accepted 15 September 2003; doi:10.1038/nature02052.
Published online 8 October 2003.

1. Jones, J. M. *et al.* mnd2: a new mouse model of inherited motor neuron disease. *Genomics* **16**, 669–677 (1993).
2. Weber, J. S. *et al.* High-resolution genetic, physical, and transcript map of the mnd2 region of mouse chromosome 6. *Genomics* **54**, 107–115 (1998).
3. Rathke-Hartlieb, S. *et al.* Progressive loss of striatal neurons causes motor dysfunction in MND2 mutant mice and is not prevented by Bcl-2. *Exp. Neurol.* **175**, 87–97 (2002).
4. Jang, W. *et al.* Comparative sequence of human and mouse BAC clones from the mnd2 region of chromosome 2p13. *Genome Res.* **9**, 53–61 (1999).
5. Jang, W., Weber, J. S., Bashir, R., Bushby, K. & Meisler, M. H. Aup1, a novel gene on mouse chromosome 6 and human chromosome 2p13. *Genomics* **36**, 366–368 (1996).
6. Jang, W., Weber, J. S., Harkins, E. B. & Meisler, M. H. Localization of the rhotekin gene RTKN on the physical maps of mouse chromosome 6 and human chromosome 2p13 and exclusion as a candidate for mnd2 and LGMD2B. *Genomics* **40**, 506–507 (1997).
7. Jang, W., Weber, J. S., Tokito, M. K., Holzbaur, E. L. & Meisler, M. H. Mouse p150Glued (dynactin 1) cDNA sequence and evaluation as a candidate for the neuromuscular disease mutation mnd2. *Biochem. Biophys. Res. Commun.* **231**, 344–347 (1997).
8. Ji, W. *et al.* DQX1, an RNA-dependent ATPase homolog with a novel DEAQ box: expression pattern and genomic sequence comparison of the human and mouse genes. *Mamm. Genome* **12**, 456–461 (2001).
9. Li, W. *et al.* Structural insights into the pro-apoptotic function of mitochondrial serine protease HtrA2/Omi. *Nature Struct. Biol.* **9**, 436–441 (2002).
10. Hegde, R. *et al.* Identification of Omi/HtrA2 as a mitochondrial apoptotic serine protease that disrupts inhibitor of apoptosis protein-caspase interaction. *J. Biol. Chem.* **277**, 432–438 (2002).
11. Suzuki, Y. *et al.* A serine protease, HtrA2, is released from the mitochondria and interacts with XIAP, inducing cell death. *Mol. Cell* **8**, 613–621 (2001).
12. Verhagen, A. M. *et al.* HtrA2 promotes cell death through its serine protease activity and its ability to antagonize inhibitor of apoptosis proteins. *J. Biol. Chem.* **277**, 445–454 (2002).
13. Martins, L. M. *et al.* The serine protease Omi/HtrA2 regulates apoptosis by binding XIAP through a reaper-like motif. *J. Biol. Chem.* **277**, 439–444 (2002).
14. van Loo, G. *et al.* The serine protease Omi/HtrA2 is released from mitochondria during apoptosis. Omi interacts with caspase-inhibitor XIAP and induces enhanced caspase activity. *Cell Death Differ.* **9**, 20–26 (2002).
15. Clausen, T., Southan, C. & Ehrmann, M. The HtrA family of proteases: implications for protein composition and cell fate. *Mol. Cell* **10**, 443–455 (2002).
16. Maurizi, M. R. Love it or leave it: tough choices in protein quality control. *Nature Struct. Biol.* **9**, 410–412 (2002).
17. Srinivasula, S. M. *et al.* Inhibitor of apoptosis proteins are substrates for the mitochondrial serine protease Omi/HtrA2. *J. Biol. Chem.* **278**, 31469–31472 (2003).
18. Yang, Q. H., Church-Hajduk, R., Ren, J., Newton, M. L. & Du, C. Omi/HtrA2 catalytic cleavage of inhibitor of apoptosis (IAP) irreversibly inactivates IAPs and facilitates caspase activity in apoptosis. *Genes Dev.* **17**, 1487–1496 (2003).
19. Brustovetsky, N. *et al.* Increased susceptibility of striatal mitochondria to calcium-induced permeability transition. *J. Neurosci.* **23**, 4858–4867 (2003).
20. Szalai, G., Krishnamurthy, R. & Hajnoczky, G. Apoptosis driven by IP(3)-linked mitochondrial calcium signals. *EMBO J.* **18**, 6349–6361 (1999).
21. Reed, J. C. Cytochrome c: can't live with it—can't live without it. *Cell* **91**, 559–562 (1997).
22. Joza, N. *et al.* Essential role of the mitochondrial apoptosis-inducing factor in programmed cell death. *Nature* **410**, 549–554 (2001).
23. Klein, J. A. *et al.* The harlequin mouse mutation downregulates apoptosis-inducing factor. *Nature* **419**, 367–374 (2002).
24. Walsh, N. P., Alba, B. M., Bose, B., Gross, C. A. & Sauer, R. T. OMP peptide signals initiate the envelope-stress response by activating DegS protease via relief of inhibition mediated by its PDZ domain. *Cell* **113**, 61–71 (2003).
25. Leist, M., Single, B., Castoldi, A. F., Kuhnle, S. & Nicotera, P. Intracellular adenosine triphosphate (ATP) concentration: a switch in the decision between apoptosis and necrosis. *J. Exp. Med.* **185**, 1481–1486 (1997).
26. Susin, S. A. *et al.* Molecular characterization of mitochondrial apoptosis-inducing factor. *Nature* **397**, 441–446 (1999).
27. Danial, N. N. *et al.* BAD and glucokinase reside in a mitochondrial complex that integrates glycolysis and apoptosis. *Nature* **424**, 952–956 (2003).
28. Casari, G. *et al.* Spastic paraplegia and OXPHOS impairment caused by mutations in paraplegin, a nuclear-encoded mitochondrial metalloprotease. *Cell* **93**, 973–983 (1998).
29. Van Dyck, L. & Langer, T. ATP-dependent proteases controlling mitochondrial function in the yeast *Saccharomyces cerevisiae*. *Cell. Mol. Life Sci.* **56**, 825–842 (1999).

Acknowledgements This work was supported by grants from the National Institutes of Health and the Muscular Dystrophy Association. S.M.S. is a Kimmel scholar. W.J. acknowledges support from the Hearing and Chemical Senses Training Program of the University of Michigan. We thank J. Zhang for help in preparation of MEFs and A. Zervos for the mouse Omi cDNA. We also thank M. Farrer and T. Gasser for the PARK3 DNA samples.

Competing interests statement The authors declare that they have no competing financial interests.

Authors' contributions J.M.J., P.D., S.M.S. and W.J. share equal first authorship and E.S.A. and M.H.M. share equal senior authorship.

Correspondence and requests for materials should be addressed to E.S.A. (E_Alnemri@lac.jci.tju.edu) or M.H.M. (meislerm@umich.edu).

Functional cloning of TUG as a regulator of GLUT4 glucose transporter trafficking

Jonathan S. Bogan^{1,2,3*}, Natalie Hendon^{1,2}, Adrienne E. McKee^{1*},
Tsu-Shuen Tsao¹ & Harvey F. Lodish^{1,4}

¹Whitehead Institute for Biomedical Research, Cambridge, Massachusetts 02142, USA

²Diabetes Unit, Department of Medicine, Massachusetts General Hospital, Boston, Massachusetts 02129, USA

³Department of Medicine, Harvard Medical School, Boston, Massachusetts 02114, USA

⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA

* Present addresses: Section of Endocrinology and Metabolism, Department of Internal Medicine, Yale University School of Medicine, 333 Cedar Street, Box 208020, New Haven, Connecticut 06520-8020, USA (J.S.B.); Program in Biological and Biomedical Sciences, Harvard Medical School, 220 Longwood Ave, Boston, Massachusetts 02115, USA (A.E.M.)

Insulin stimulates glucose uptake in fat and muscle by mobilizing the GLUT4 glucose transporter. GLUT4 is sequestered intracellularly in the absence of insulin, and is redistributed to the plasma membrane within minutes of insulin stimulation^{1,2}. But the trafficking mechanisms that control GLUT4 sequestration have remained elusive. Here we describe a functional screen to identify proteins that modulate GLUT4 distribution, and identify TUG as a putative tether, containing a UBX domain, for GLUT4. In truncated form, TUG acts in a dominant-negative manner to inhibit insulin-stimulated GLUT4 redistribution in Chinese hamster ovary cells and 3T3-L1 adipocytes. Full-length TUG forms a complex specifically with GLUT4; in 3T3-L1 adipocytes, this complex is present in unstimulated cells and is largely disassembled by insulin. Endogenous TUG is localized with the insulin-mobilizable pool of GLUT4 in unstimulated 3T3-L1 adipocytes, and is not mobilized to the plasma membrane by insulin. Distinct regions of TUG are required to bind GLUT4 and to retain GLUT4 intracellularly in transfected, non-adipose cells. Our data suggest that TUG traps endocytosed GLUT4 and tethers it intracellularly, and that insulin mobilizes this pool of retained GLUT4 by releasing this tether.

The proportion of GLUT4 at the surface of individual living cells was assayed by using flow cytometry to measure fluorescence intensities corresponding to cell surface and total GLUT4. For this, we used a previously described GLUT4 reporter protein containing epitope tags in its first extracellular domain and green fluorescent protein (GFP) at the carboxy terminus³. To identify proteins that alter GLUT4 distribution, we constructed a 3T3-L1 adipocyte complementary DNA library in a retroviral vector. We used the pMX-IRES-CD2 vector because it produces a bicistronic

message encoding both the cloned cDNA and the extracellular portion of CD2 antigen, which makes it easier to identify infected cells⁴. We infected a clonal Chinese hamster ovary (CHO) cell line expressing the GLUT4 reporter and the murine ecotropic retrovirus receptor, such that 34% of $\sim 5 \times 10^7$ target cells became CD2 positive. At this viral titre, statistics suggest that almost all of the infected cells received only one retrovirus from the library. Expression of each cDNA is stable because the retrovirus integrates into the genome. We enriched particular cDNAs within the library that, when overexpressed, alter trafficking of the GLUT4 reporter.

Our previous data support the notion that, in CHO cells cultured with abundant amino acids, insulin mobilizes GLUT4 to the plasma membrane from endosomes and a poorly defined, highly insulin-responsive compartment that may be similar to that found in 3T3-L1 adipocytes³. We considered that 5 min after insulin addition, but not necessarily after more prolonged stimulation, most of the increased GLUT4 at the cell surface will come from the highly insulin-responsive compartment. Thus, to bias our cloning strategy towards proteins involved in trafficking through this compartment,

we cultured CHO cells containing the cDNA library in an abundance of amino acids and stimulated cells with insulin for 5 min only. We then chilled the cells and prepared them for flow cytometry. To identify cells that had either enhanced or diminished translocation of GLUT4, we used an iterative enrichment process in which we first sorted the $\sim 0.2\%$ of cells with the highest or lowest ratio of fluorescence intensities. We expanded the sorted cells in culture, then restimulated the expanded cells with insulin and subjected them to further enrichment by flow sorting. Figure 1a illustrates this approach to enriching cells with diminished translocation of GLUT4. For the third flow cytometry step, our sorting criteria were less stringent, and we cloned individual cells in 96-well plates if they fell within the $\sim 1\%$ of the population with the highest or lowest ratio of fluorescence intensities. We predicted that after three iterations of sorting and expansion, the overall enrichment would be $>10^7$ -fold, which is sufficient to isolate individual cDNAs from the library. Of the 150 clonal cell lines isolated, 77% were CD2 positive. Most (142/150) of these cell lines had increased GLUT4 at the cell surface, because cloning efficiency was poor for cells with

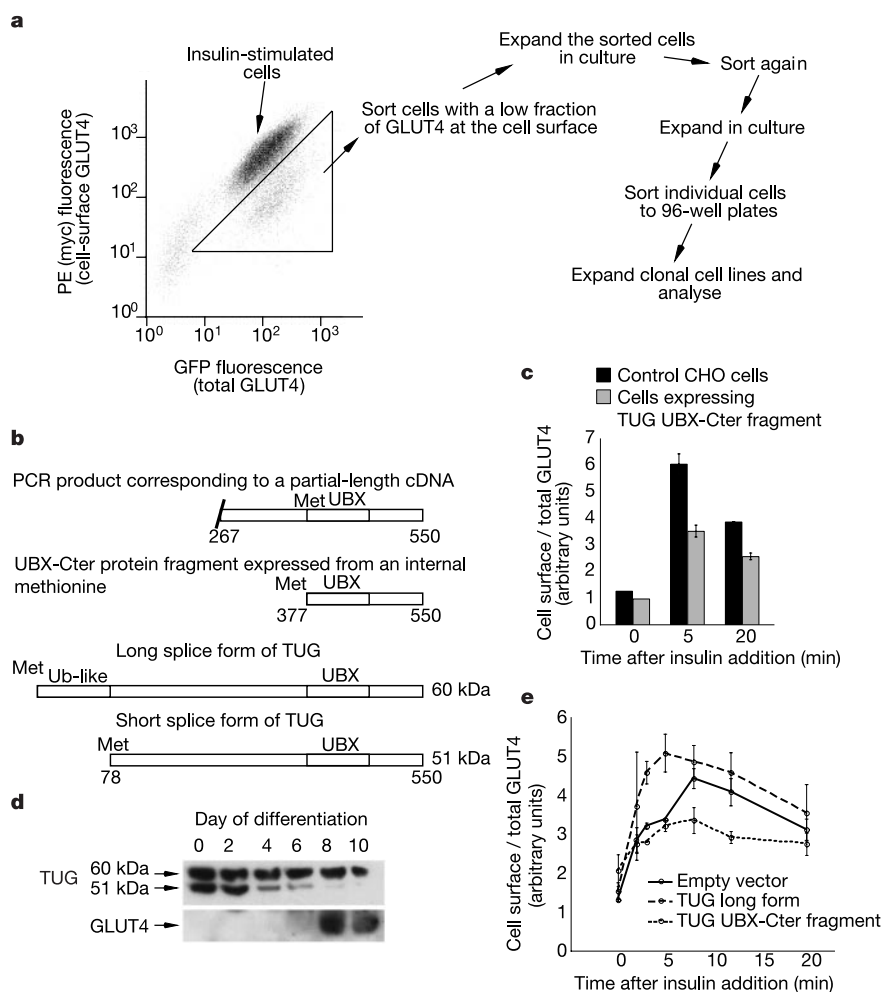


Figure 1 Expression cloning of TUG. **a**, Enrichment process in cells with a low ratio of surface to total GLUT4. See main text and Methods for more details. Note that the upper diagonal edge of the population (high surface:total GLUT4) is sharper than the lower diagonal edge of the population (low surface:total GLUT4) owing to fewer dead or poorly stained cells, permitting a more stringent sorting gate. **b**, PCR of proviral insertions from genomic DNA of two clones yielded identical 1.2-kb products. This sequence contained a partial-length open reading frame, which spans TUG residues 267–550 and contains a UBX domain. Translation beginning at the first methionine is predicted to produce a

protein fragment known as UBX-Cter (residues 377–550). The full-length TUG protein is present in two splice forms. The longer of these forms contains a region at its N terminus with similarity to ubiquitin. **c**, Expression of TUG UBX-Cter inhibits the ability of insulin to redistribute GLUT4 in CHO cells. **d**, The short form of TUG is downregulated during adipose differentiation, whereas the long form persists and is expressed together with GLUT4 in differentiated adipocytes. **e**, Opposite effects of expression of the long form of TUG and of TUG UBX-Cter on GLUT4 trafficking in 3T3-L1 adipocytes. Measurements were made in triplicate, and error bars indicate standard deviation.

decreased GLUT4 translocation. We prepared genomic DNA from selected CD2-positive cell lines with consistently altered GLUT4 trafficking, and used polymerase chain reaction (PCR) with primers from sequences in the retroviral vector to identify the inserted cDNAs.

Two out of the eight clones isolated in the screen for cells with a low fraction of GLUT4 at the cell surface had a particularly robust phenotype, and PCR of genomic DNA from these clones yielded identical 1.2-kilobase (kb) products. Sequencing identified an 852-nucleotide reading frame encoding the C terminus of a novel protein, which we call TUG (tether, containing a UBX domain, for GLUT4) (Fig. 1b). We reasoned that the cloned sequence would produce a protein fragment initiated at the first encoded methionine; this fragment would correspond to the C terminus of full-length TUG. Searches of the murine expressed sequence tag (EST) database identified two splice forms of TUG. The longer of these contains a sequence at the amino terminus with similarity to ubiquitin, and is predicted to be 550 residues long. The probable start codon in the short form is equivalent to methionine 78 of the long form. By this numbering, the C-terminal fragment might start at methionine 377, which is normally an internal residue. Database searches identified a 79-residue UBX domain in the predicted TUG protein. Other proteins containing this domain include p47, an adaptor protein that recruits the NSF (*N*-ethylmaleimide-sensitive factor)-like p97 ATPase in the fusion of Golgi and transitional endoplasmic reticulum (ER) membranes with FAF1 (Fas-associated factor 1), a protein that binds to Fas and regulates apoptosis^{3,6}. The structures of the UBX domains in p47 and FAF1 indicate that they form a ubiquitin-like, β -grasp structure^{7,8}.

The C-terminal TUG fragment we predicted to be produced by our initial clone contains this UBX domain (Fig. 1b), and we therefore refer to this fragment as TUG UBX-Cter. We further considered that production of UBX-Cter might alter the insulin-stimulated translocation of GLUT4. Consistent with this hypoth-

esis, expression of UBX-Cter substantially inhibits the ability of insulin to redistribute GLUT4 in CHO cells (Fig. 1c). These results show that our cloning strategy was successful and, together with the data presented below, indicate that this fragment acts in a dominant-negative manner to block the insulin signal required to redistribute GLUT4.

During our study, the probable human homologue of TUG was identified as part of a fusion protein found in alveolar soft-part sarcomas⁹. This protein, termed ASPL, is 76% identical to murine TUG and also undergoes alternative splicing. Both TUG and ASPL are widely expressed⁹ (Supplementary Fig. 1). Antisera raised against the C terminus of TUG/ASPL were specific when used for immunoblotting or immunofluorescence microscopy (Supplementary Fig. 2). In 3T3-L1 cells, immunoblotting shows that the short form of TUG is downregulated during adipose differentiation, whereas the long form persists and is expressed together with GLUT4 in differentiated adipocytes (Fig. 1d). We conclude, from this and the data below, that it is the long form of TUG that controls GLUT4 distribution in 3T3-L1 adipocytes.

To determine whether the C-terminal fragment of TUG inhibits insulin-stimulated GLUT4 translocation in 3T3-L1 adipocytes as well as in CHO cells, we expressed UBX-Cter by retrovirally infecting cells expressing the GLUT4 reporter. As shown in Fig. 1e, expression of this truncated protein substantially inhibits the degree to which insulin increases GLUT4 at the plasma membrane. Some of this effect is due to increased GLUT4 at the plasma membrane in the absence of insulin; this increase was not observed in CHO cells, which are a poorer adipocyte model. Overexpression of the full-length TUG long form in 3T3-L1 adipocytes increases the rapidity and initial magnitude of GLUT4 movement to the surface after insulin stimulation. Thus, truncated and full-length forms of TUG have opposite effects on insulin-stimulated movement of GLUT4 when they are expressed exogenously in 3T3-L1 adipocytes. The full-length TUG protein enhances the normal action of insulin,

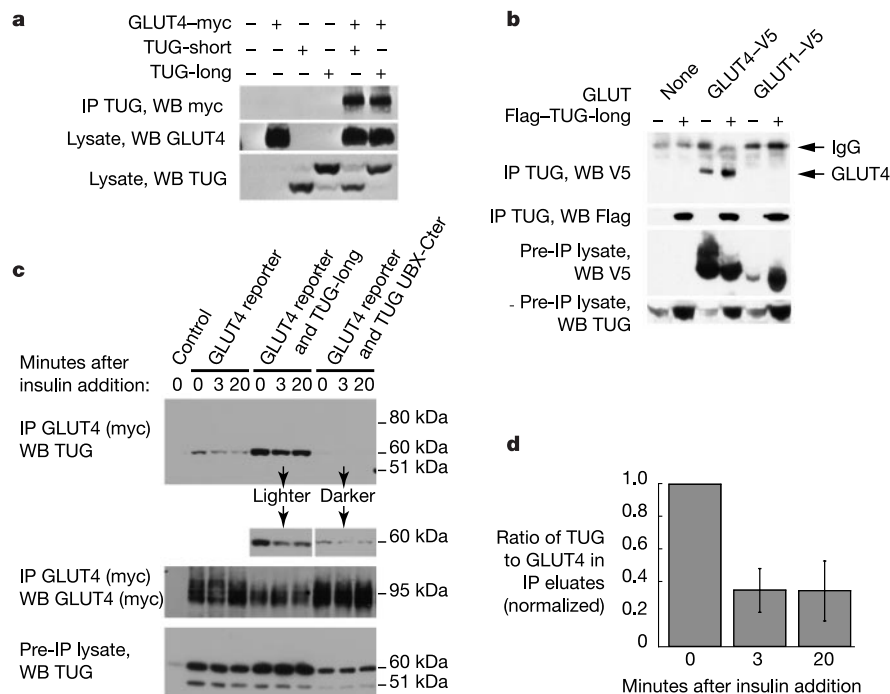


Figure 2 Interaction of TUG with GLUT4. **a**, Immunoprecipitation of either the long or short form of TUG co-purifies myc-tagged GLUT4 in transfected 293T cells. **b**, The interaction of TUG with GLUT4 is specific, because GLUT1 containing an identical epitope tag is not co-immunoprecipitated. **c**, Interaction of endogenous TUG with GLUT4 in 3T3-L1 adipocytes. Insulin decreases the amount of TUG immunoprecipitated with GLUT4.

Expression of the TUG long form increases the amount of TUG that is co-purified with GLUT4, and expression of TUG UBX-Cter decreases this amount. **d**, Insulin decreases the amount of TUG complexed with GLUT4. Measurements were in quadruplicate, and error bars indicate standard deviation.

whereas the UBX-Cter fragment acts in a dominant-negative fashion to inhibit the insulin-stimulated movement of GLUT4. As detailed below, we believe that overexpression of TUG increases the fraction of GLUT4 that accumulates in an intracellular, insulin-mobilizable pool in unstimulated cells.

To determine whether the effects of truncated and full-length TUG overexpression in 3T3-L1 adipocytes are specific to GLUT4, we immunoblotted plasma-membrane fractions to detect endogenous GLUT4, GLUT1 and the transferrin receptor (TfnR). We treated cells briefly with insulin to detect the peak of the translocation response. As shown in Supplementary Fig. 3, insulin markedly increases GLUT4 in plasma membranes of control cells, and this effect is enhanced by overexpression of full-length TUG and is blocked by expression of the UBX-Cter fragment. These data are consistent with the results shown in Fig. 1e. Insulin slightly increases the amounts of GLUT1 and TfnR in plasma membranes of control cells. Exogenous expression of full-length TUG or UBX-Cter fragment has little effect on the insulin-stimulated increase in plasma-membrane GLUT1 and TfnR. We immunoblotted the insulin receptor β -subunit as a negative control; as expected, insulin did not alter the amount of β -subunit present in the plasma membranes.

We also examined whether TUG influences the ability of insulin to regulate 2-deoxyglucose uptake in 3T3-L1 adipocytes. We found

that insulin stimulates glucose uptake by 3.6-fold in control cells (95% confidence interval, 1.6-fold to 8.3-fold), by 3.5-fold in cells stably expressing the long form of TUG (1.9-fold to 6.3-fold) and by 0.73-fold in cells stably expressing the TUG UBX-Cter fragment (0.54-fold to 1.0-fold). Total amounts of GLUT4 and GLUT1 were similar in all samples as judged by immunoblotting, and exogenous expression of TUG or of the UBX-Cter fragment had no effect on 3T3-L1 differentiation (data not shown). Thus, TUG UBX-Cter blocks the ability of insulin to stimulate glucose uptake, and this effect cannot be attributed to altered glucose transporter levels. Taken together, the data indicate that TUG specifically links insulin action with GLUT4 movement, and its action modulates deoxyglucose uptake in 3T3-L1 adipocytes.

To gain insight into how a UBX-domain-containing protein controls GLUT4 trafficking, we reviewed the functions of other proteins containing this domain. One such protein, FAF1, binds to wild-type Fas but not to the allele of Fas found in *lpr*^{cg} mice, which has a single point mutation in its cytoplasmic domain⁶. Comparison of the sequences of GLUT4 and Fas revealed a small region of local similarity (236-EAKKFARENNI-246 in Fas and 263-EKRKLER-ERPL-273 in GLUT4) within the large cytoplasmic loop between the sixth and seventh transmembrane domains of GLUT4. More significantly, the I246N mutation in the Fas *lpr*^{cg} allele falls within this motif and corresponds to a similar residue (L273) in GLUT4. In

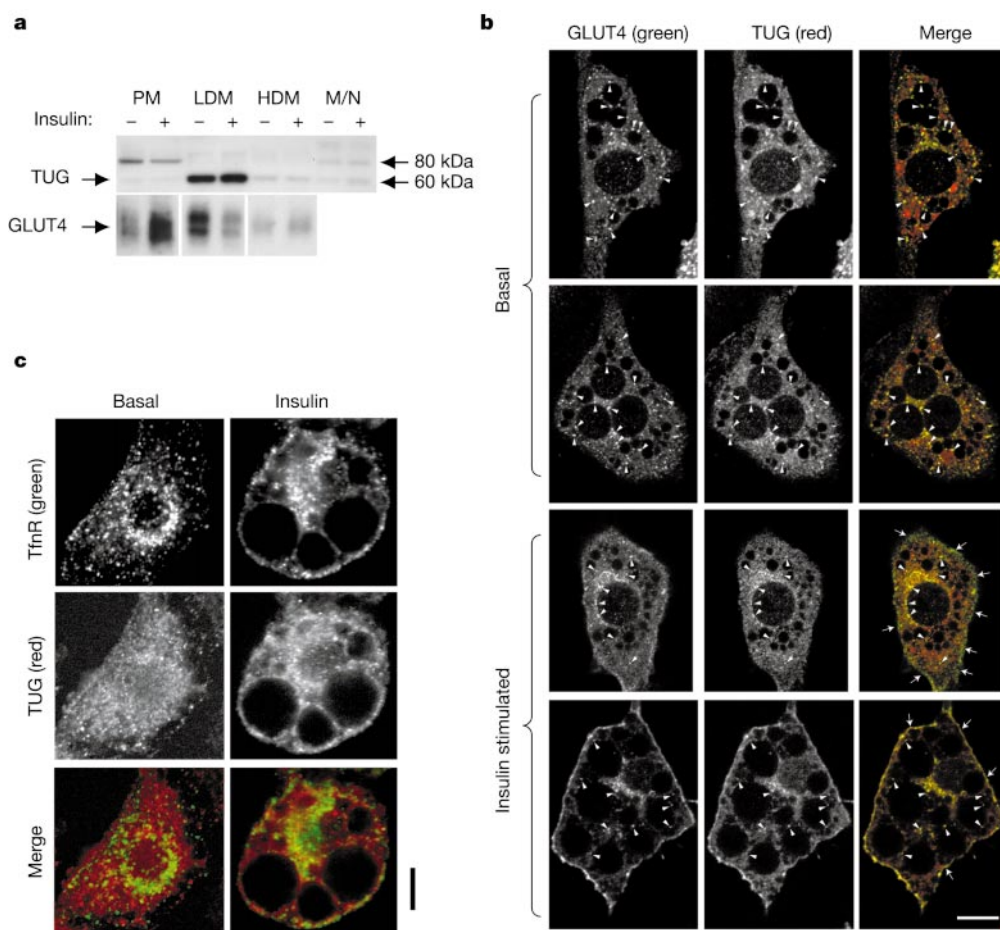


Figure 3 Subcellular localization of TUG. **a**, Basal and stimulated 3T3-L1 adipocytes were subjected to differential centrifugation to isolate fractions enriched in plasma membranes (PM), low-density microsomes (LDM), high-density microsomes (HDM) and mitochondria and nuclei (M/N). Equal amounts of protein from each fraction were immunoblotted with anti-TUG and anti-GLUT4 antibodies. **b**, Basal and stimulated 3T3-L1 adipocytes expressing the GLUT4 reporter were stained to detect anti-GLUT4 (green) and anti-TUG

(red) antibodies. Images were acquired by confocal microscopy. Arrowheads indicate colocalization. Arrows show regions of plasma-membrane staining for GLUT4 but not TUG. Bar, 10 μ M. **c**, Basal and stimulated 3T3-L1 adipocytes were stained using anti-TfnR (green) and anti-TUG (red) antibodies. Images were acquired by confocal microscopy. Bar, 10 μ M.

addition, Ubc9, a conjugating enzyme for the ubiquitin-like protein Sumo, binds to Fas and GLUT4 (refs 10–12). Daxx is another protein recently reported to interact with GLUT4 as well as with Fas¹³. Therefore, we proposed that similar complexes involving Ubc9, Daxx and a UBX-containing protein form with Fas and GLUT4. To test the prediction that GLUT4 and TUG interact, we performed co-immunoprecipitation experiments using transiently transfected 293T cells. Immunoprecipitation of either the long or the short form of TUG co-purifies myc-tagged GLUT4 (Fig. 2a). This interaction is specific for GLUT4, because GLUT1 is not co-purified in similar co-transfection experiments (Fig. 2b); some GLUT4 is even immunoprecipitated with endogenous TUG. These data are consistent with those in Supplementary Fig. 3, which implicate TUG in GLUT4 but not GLUT1 trafficking. Thus, TUG complexes interact specifically with GLUT4 in transfected cells, and the N-terminal ubiquitin-like domain of TUG is not required for this interaction.

To determine whether endogenous TUG interacts with GLUT4 in 3T3-L1 adipocytes, we used cells expressing the GLUT4 reporter; this tagged protein is expressed at approximately fivefold higher levels than endogenous GLUT4 in mature 3T3-L1 adipocytes (data not shown). We stimulated cells with insulin for various times, then prepared lysates and immunoprecipitated the reporter using an anti-myc affinity matrix. TUG was present in the eluted material (Fig. 2c), which further demonstrates that insulin decreases the amount of TUG that immunoprecipitates with GLUT4. Densitometric analysis indicates that insulin reduces by two-thirds the amount of TUG that is complexed with GLUT4 (Fig. 2d). In cells overexpressing the long form of TUG, which accelerates the mobilization of GLUT4 after insulin addition, more TUG immunoprecipitates with GLUT4. Conversely, in cells expressing TUG UBX-Cter, less TUG is co-immunoprecipitated. In all cases, insulin

decreases the association between TUG and GLUT4. Thus, TUG and GLUT4 form complexes in 3T3-L1 adipocytes, and insulin stimulates disassembly of most of these complexes.

To assess the subcellular distribution of endogenous TUG, we fractionated 3T3-L1 adipocytes using a standard differential centrifugation protocol. TUG is present in approximately equal amounts in the cytosol and in membranes (data not shown); membrane-associated TUG is present predominantly in the low-density microsomal fraction (Fig. 3a); the same fraction contains the highly insulin-responsive pool of GLUT4. In contrast to GLUT4, we observed no marked movement of TUG after insulin stimulation. In addition to the expected 60-kDa band corresponding to the long form of TUG, we observe an 80-kDa band in the plasma-membrane fraction. This 80-kDa band may be a cross-reacting protein. Another possibility is that it is a form of TUG that is modified by covalent addition of a ubiquitin-like protein, because our antisera seem to be specific (Supplementary Fig. 2a) and the Sumo-conjugating protein Ubc9 also binds to GLUT4 (ref. 10). The 80-kDa protein is not present in material immunoprecipitated with GLUT4, and therefore does not form a stable complex with GLUT4 (Fig. 2c).

To test the colocalization of TUG and GLUT4, we performed confocal immunofluorescence microscopy of 3T3-L1 adipocytes expressing the GLUT4 reporter. As shown in Fig. 3b, TUG has a homogeneous distribution, corresponding to cytosolic protein, and an overlying punctate and perinuclear distribution, corresponding to membrane-associated TUG. In unstimulated cells, most TUG punctae localize with GLUT4 (arrowheads), which is present in its characteristic vesicular and perinuclear pattern. After insulin addition, GLUT4 is clearly visible at the plasma membrane, and there seem to be fewer intracellular punctae where GLUT4 and TUG overlap. TUG is also present at or near the plasma membrane,

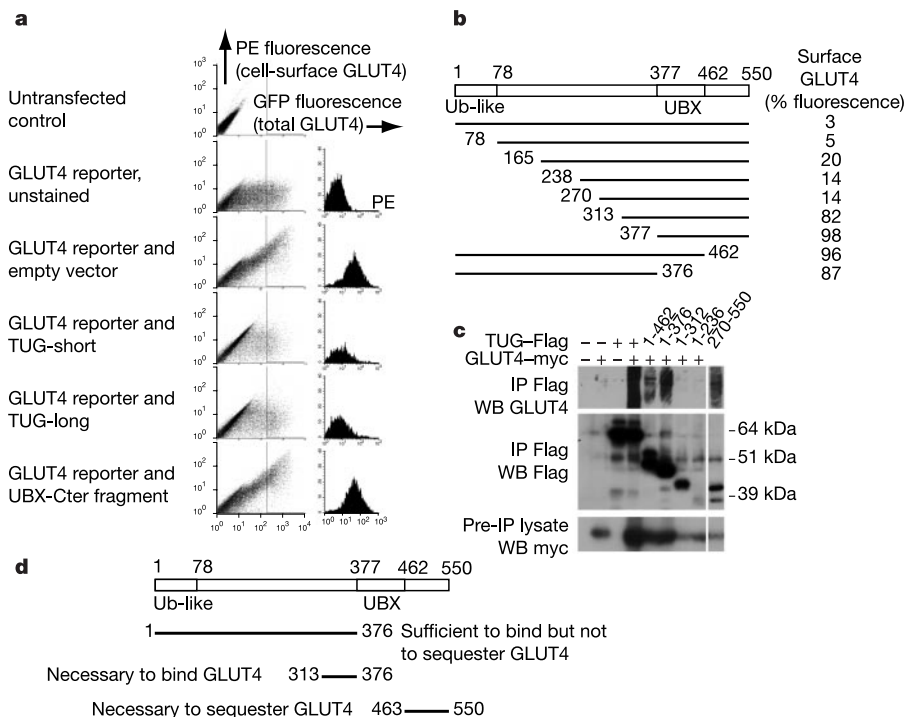


Figure 4 Mapping of TUG domains in non-adipose cells. **a**, Expression of either the long or the short form of TUG reduces the appearance of GLUT4 at the surface, as detected by flow cytometry. The histograms to the right indicate the distribution of phycoerythrin (PE), which corresponds to cell-surface GLUT4. **b**, Deletion of the N terminus to residue 165 slightly reduces retention of GLUT4, as assessed by PE fluorescence. Deletions beyond residue 270 prevent TUG from sequestering GLUT4. Deletion of the C terminus also

prevents sequestration. **c**, Truncation of C-terminal residues 377–550 and 463–550 does not impair immunoprecipitation of GLUT4. N-terminal truncations of residues 1–269 do not impair immunoprecipitation either. **d**, A summary of the regions of TUG that are (1) sufficient to bind GLUT4 in immunoprecipitation experiments, (2) necessary for this interaction, and (3) necessary to sequester GLUT4 in transfected cells but not required to form a complex with GLUT4.

although to a lesser extent than GLUT4 in insulin-stimulated cells; there are clear regions at the cell surface that stain for GLUT4 but not TUG (arrows). The TUG observed at the plasma membrane may correspond to the 80-kDa band seen by subcellular fractionation (Fig. 3a). To determine whether the punctae labelled with the TUG antibody correspond to endosomes, we acquired confocal images of 3T3-L1 adipocytes stained to detect endogenous TUG and TfnR. These proteins have distinct and non-overlapping distributions (Fig. 3c). Thus, TUG and GLUT4 colocalize on intracellular, insulin-responsive membranes that are distinct from TfnR-containing endosomes, and TUG may also be present at or near the plasma membrane in a covalently modified form.

Our data suggest that TUG is an essential part of the mechanism by which GLUT4 is sequestered in unstimulated cells. In this model, intracellular GLUT4 is trapped by a complex comprising TUG and probably other proteins, which stably associates with GLUT4 and retains it in intracellular, non-endosomal membranes by binding to an unknown anchoring protein. To test directly whether TUG sequesters GLUT4 in non-adipose cells, we performed flow cytometry of 293T cells transiently transfected with the GLUT4 reporter with or without TUG. Co-expression of either form of TUG greatly reduces the appearance of GLUT4 at the surface (Fig. 4a). In these co-transfections, the fluorescent signal corresponding to cell-surface GLUT4 is only 3–5% of that in control cells not transfected with TUG. These data are consistent with those in Fig. 1e and Supplementary Fig. 3, in which overexpression of the full-length form of TUG apparently increased the size of the rapidly insulin-mobilizable intracellular pool of GLUT4 in unstimulated 3T3-L1 adipocytes. Co-transfection of the truncated TUG UBX-Cter fragment does not cause intracellular retention of GLUT4 in 293T cells, consistent with the idea that this fragment blocks formation of a complex required for GLUT4 retention.

To map the regions of TUG required to sequester GLUT4, we co-transfected 293T cells with various truncations of TUG and the GLUT4 reporter, and, similar to Fig. 4a, assayed cell-surface GLUT4 by flow cytometry. Deletion of the N terminus to residue 165 modestly decreases the ability of TUG to retain GLUT4 intracellularly (Fig. 4b). Further deletions of the N terminus, from residue 270 to 313 or beyond, prevent TUG from sequestering GLUT4, as does deletion of residues to 462 or beyond from the C terminus of TUG. In complementary experiments, we sought to determine by co-immunoprecipitation which regions of TUG complex with GLUT4. Truncation of the C terminus of TUG (residues 377–550 or 463–550) does not impair its ability to co-immunoprecipitate GLUT4 (Fig. 4c). Likewise, an N-terminal truncation (of residues 1–269) still co-immunoprecipitates GLUT4. Thus, residues 1–376 of TUG are sufficient to complex with GLUT4, but not sufficient to retain it in non-adipose cells. Within this region, residues 313–376 are necessary to form the complex: truncation of these residues abrogates GLUT4 co-immunoprecipitation. Finally, the C terminus of TUG (residues 463–550) is required to retain GLUT4 intracellularly within non-adipose cells but not for the interaction of TUG with GLUT4. This region may bind an intracellular anchor to retain the TUG–GLUT4 complex within cells.

Insulin regulates glucose homeostasis on a minute-to-minute basis by controlling glucose use by muscle and fat^{1,2,14}. Trafficking of GLUT4 is crucial for this regulation, and insulin probably acts at several steps in the GLUT4 recycling pathway to control its subcellular distribution. After endocytosis from the plasma membrane, GLUT4 is sorted away from endosomes to a highly insulin-responsive compartment, where it is selectively retained. Some data suggest that endocytosed GLUT4 moves through the trans-Golgi network (TGN); if so, this step might precede its arrival in the insulin-responsive compartment, or the insulin-responsive pool might participate in an intracellular cycle through the TGN^{1,15}. Our data are compatible with these models, and TUG could retain GLUT4 in post-TGN membranes or constrain a dynamic intracellular cycle.

Most biochemical and kinetic data are consistent with a three-compartment model in which GLUT4 is present predominantly in the plasma membrane, endosomes/TGN and insulin-responsive membranes^{2,3}. In this model, TUG binds endocytosed GLUT4 through its N-terminal and central domains and retains it in post-endocytic, insulin-responsive membranes through its C terminus; presumably, the C terminus binds an unidentified intracellular anchor. Insulin mobilizes this pool of retained GLUT4 by disrupting the interaction between TUG and GLUT4. Thus, the data suggest that in adipocytes the long form of TUG tethers a highly insulin-responsive pool of GLUT4.

The effects of TUG overexpression on the transient characteristics of insulin-stimulated GLUT4 translocation are consistent with the above interpretation. Our previous data show that insulin-stimulated movement of GLUT4 is characterized by a brief overshoot of the steady-state distribution³. This response is cell-type-specific, and is consistent with a three-compartment model for GLUT4 trafficking. GLUT4 mobilized from the insulin-responsive, non-endosomal compartment accounts for the overshoot. It is significant that overexpression of TUG increases the rapidity and initial magnitude of GLUT4 translocation, which is exactly what would be predicted if the size of the insulin-mobilized pool were increased. Therefore, the data in Fig. 1e and Supplementary Fig. 3 provide an essential functional corroboration of the binding and colocalization data in Figs 2 and 3. Moreover, confocal microscopy indicates that most intracellular, non-endosomal GLUT4 localizes with TUG, consistent with the idea that this represents a relatively stable pool in the absence of insulin.

We do not yet know the insulin signaling pathway(s) that causes dissociation of TUG and GLUT4. This dissociation occurs soon after insulin addition and precedes significant movement of GLUT4 (ref. 3). Insulin signals through at least two distinct mechanisms to redistribute GLUT4; these involve phosphatidylinositol-3-kinase activation, which signals to Akt and perhaps to other proteins, as well as a CAP–Cbl–TC10 pathway leading to the mammalian exocyst complex and probably to other effectors^{16,17}. These pathways may act on different steps in the GLUT4 recycling pathway; for example, one signal may move GLUT4 out of an intracellular, insulin-responsive compartment, and another may stimulate fusion of vesicles containing GLUT4 at the plasma membrane². We find that the TUG UBX-Cter fragment disrupts the interaction of GLUT4 and native TUG; unstimulated cells expressing this fragment have slightly increased GLUT4 at the plasma membrane, and insulin-induced GLUT4 translocation is markedly inhibited. Expression of the UBX-Cter fragment probably prevents GLUT4 from being retained in an insulin-responsive pool but is not sufficient to bypass other regulated steps in the recycling pathway. GLUT4 might pass through the insulin-regulated compartment to be loaded onto molecular motors containing the unconventional myosin Myo1c, which could maintain a rapid exocytosis rate (compared with endosome recycling) in the presence of insulin¹⁸.

The observation that some TUG is at or near the plasma membrane of cells treated with insulin raises the possibility that TUG positions GLUT4 near the plasma membrane after insulin stimulation. It is interesting that the TUG antisera detect an 80-kDa band in the plasma membrane fraction. Given the likely presence of Ubc9 in the complex, it may be that this band represents TUG that has been modified with a ubiquitin-like protein. Sumo and other ubiquitin-like modifications regulate the intracellular targeting of proteins¹⁹. Insulin might stimulate the covalent modification of TUG; this may be transient but nonetheless sufficient to target it, together with GLUT4, to the plasma membrane.

Our data show that both forms of TUG can bind and sequester GLUT4 in transfected cells, yet only the long form is expressed in mature 3T3-L1 adipocytes. One possibility is that the relative amounts of long and short TUG forms control the ability of insulin to mobilize tethered GLUT4. For example, it may be that only

GLUT4 that is retained by the long form of TUG is mobilized by insulin, and that the short form of TUG tethers GLUT4 in a non-insulin-responsive pool. If so, the ability of insulin to control glucose uptake might be controlled at the level of TUG transcription. We do not know whether the long or short splice forms of TUG are targets of insulin action in muscle, nor whether TUG has any role in contraction-mediated GLUT4 translocation. Nonetheless, TUG may have an important role in controlling organism-level glucose homeostasis, and variations in this protein may predispose humans to insulin resistance characteristic of type 2 diabetes mellitus. □

Methods

Library construction, retroviral infection and cell culture

A cDNA library was constructed from 3T3-L1 adipocyte poly-A + RNA using a combination of oligo-dT and random hexamer priming for first-strand synthesis (Superscript, Life Technologies). cDNAs were cloned non-directionally in the *Bst*XI sites of the pMX-IRES-CD2 vector⁴. The total complexity of the library was calculated at $>2.4 \times 10^6$ independent clones, with $>90\%$ of clones having inserts. The retrovirus was produced by transient transfection of 293-based packaging lines as described previously³, except that pCL-Eco was co-transfected to increase viral titres²⁰. Virus was used to infect CHO cells expressing the GLUT4 reporter as described previously³. For experiments in which cloned forms of the TUG cDNA were expressed from a retrovirus, the pBICD2 vector was used. pBICD2 was identical to pMX-IRES-CD2 except that point mutations were introduced to eliminate two potential translation start sites 5' of the cloning site (J.S.B., A.E.M., X. Liu and H.F.L., unpublished data). Cell culture and flow cytometry, including measurement of GLUT4 trafficking, were performed as described previously³.

Cloning and database searches

Genomic DNA was prepared from clonal cell lines using a Qiagen kit, and proviral insertions were amplified by PCR primed from retroviral vector sequences. The full-length TUG short form cDNA was cloned from an arrayed liver cDNA library (Origene). Additional sequence for the full-length TUG long-form cDNA was identified using overlapping ESTs, and this sequence was PCR-amplified from skeletal muscle cDNA (Invitrogen). Truncated cDNAs were cloned using pcDNA3.1-TOPO (Invitrogen), and all constructs were verified by DNA sequencing. Database searches used the SMART (<http://smart.embl-heidelberg.de/>) and Pfam (<http://www.sanger.ac.uk/Software/Pfam/>) databases. CD-Blast searches identify the UBQ domain with an $E = 2 \times 10^{-5}$ and the N-terminal ubiquitin-like domain in the long form with $E = 0.046$.

Antisera and biochemistry

Antisera were raised in rabbits using a peptide corresponding to the TUG C terminus, CSLGKVPKWLKLPASKR, conjugated to KLH (Covance). TUG antisera were used at dilutions of 1:1,000–2,000 on western blots. Other antibodies used included those directed to the myc epitope (9E10, Roche), Flag epitope (M2, Sigma), V5 epitope (Invitrogen), GLUT4 C terminus (gift of M. Charron), GLUT1 (Chemicon), insulin receptor β -subunit (Transduction Labs) and transferrin receptor (Pharmingen and Santa Cruz Biotechnology). Phycoerythrin-conjugated secondary antibodies were from Jackson ImmunoResearch. Human GLUT4 and GLUT1 clones tagged identically with the V5 epitope at their C terminus were from Invitrogen. 293T cells were transfected using Eugene 6 reagent (Roche). Immunoprecipitations were done in TNET or HNET (20 mM Tris pH 8.0 or 40 mM HEPES pH 7.5, with 150 mM NaCl, 2 mM EDTA, 1% Triton X-100) containing one Complete protease inhibitor tablet (Roche) per 20 ml, 20 mM iodoacetamide (Sigma), and (for 3T3-L1 cells) 50 mM octylglucoside (Roche) and 5 $\mu\text{g ml}^{-1}$ MG-132 (Calbiochem). Cells were lysed for >30 min before pelleting of insoluble debris, and immunoprecipitations were allowed to proceed for 6 h to overnight at 4 °C. Immunoprecipitates were washed six times with the same buffer, and were eluted in SDS-PAGE (polyacrylamide gel electrophoresis) sample buffer for 10–15 min at 42 °C. Western blotting and subcellular fractionation were performed as described previously³.

2-Deoxyglucose uptake

Assays were modified from ref. 21. 3T3-L1 adipocytes were seeded to 24-well plates, starved, washed with Krebs–Ringer phosphate buffer²¹, and stimulated using 100 nM insulin for 20 min at 37 °C. 1 $\mu\text{Ci ml}^{-1}$ 2-deoxy-D-[2,6-³H]glucose and 0.2 $\mu\text{Ci ml}^{-1}$ D-[1-¹⁴C]mannitol were added so that the final concentration of 2-deoxyglucose was 50 μM . Uptake was at 37 °C for 6 min, then cells were washed with cold PBS and lysed in 300 μl per well TNET. [¹⁴C]mannitol counts were used to control for residual extracellular 2-deoxyglucose, and nonspecific uptake was measured in the presence of 20 μM cytochalasin B. A portion of each lysate was used in a micro-BCA protein assay (Pierce), and counts were normalized to this measurement to control for variation in the number of cells in each well. Lysates were also immunoblotted to demonstrate equivalent expression of GLUT1 and GLUT4. Assays were done in quadruplicate. All cells contained the dual-tagged GLUT4 reporter, and control experiments showed that the presence of this protein did not contribute to glucose transport activity. Confidence intervals for the ratio of insulin-stimulated to basal deoxyglucose uptake were calculated using a *t*-test on log-transformed data.

Microscopy

3T3-L1 adipocytes expressing the GLUT4 reporter were reseeded to coverslips and used for microscopy as described previously²². Polyclonal anti-TUG antisera (1:250) and either monoclonal anti-myc (1:250) or anti-TfnR (5 $\mu\text{g per ml}$) antibodies were used with Alexa594-conjugated goat anti-rabbit IgG and Alexa488-conjugated goat anti-mouse or anti-rat IgG secondary antibodies (Molecular Probes). The myc epitope was stained to increase the signal over that due to GFP alone. Control images were acquired with each antibody individually and in the absence of primary antibody. Images were acquired on a Zeiss LSM 510 laser scanning confocal microscope.

Received 17 June; accepted 4 August 2003; doi:10.1038/nature01989.

- Bryant, N. J., Govers, R. & James, D. E. Regulated transport of the glucose transporter GLUT4. *Nature Rev. Mol. Cell Biol.* **3**, 267–277 (2002).
- Holman, G. D. & Sandoval, I. V. Moving the insulin-regulated glucose transporter GLUT4 into and out of storage. *Trends Cell Biol.* **11**, 173–179 (2001).
- Bogan, J. S., McKee, A. E. & Lodish, H. F. Insulin-responsive compartments containing GLUT4 in 3T3-L1 and CHO cells: regulation by amino acid concentrations. *Mol. Cell. Biol.* **21**, 4785–4806 (2001).
- Liu, X. *et al.* Generation of mammalian cells stably expressing multiple genes at predetermined levels. *Anal. Biochem.* **280**, 20–28 (2000).
- Kondo, H. *et al.* p47 is a cofactor for p97-mediated membrane fusion. *Nature* **388**, 75–78 (1997).
- Chu, K., Niu, X. & Williams, L. T. A Fas-associated protein factor, FAF1, potentiates Fas-mediated apoptosis. *Proc. Natl Acad. Sci. USA* **92**, 11894–11898 (1995).
- Buchberger, A., Howard, M. J., Proctor, M. & Blycroft, M. The UBQ domain: a widespread ubiquitin-like module. *J. Mol. Biol.* **307**, 17–24 (2001).
- Yuan, X. *et al.* Solution structure and interaction surface of the C-terminal domain from p47: a major p97-cofactor involved in SNARE disassembly. *J. Mol. Biol.* **311**, 255–263 (2001).
- Ladanyi, M. *et al.* The der(17)t(X;17)(p11;q25) of human alveolar soft part sarcoma fuses the TFE3 transcription factor gene to ASPL, a novel gene at 17q25. *Oncogene* **20**, 48–57 (2001).
- Giorgino, F. *et al.* The sentrin-conjugating enzyme mUbc9 interacts with GLUT4 and GLUT1 glucose transporters and regulates transporter levels in skeletal muscle cells. *Proc. Natl Acad. Sci. USA* **97**, 1125–1130 (2000).
- Becker, K., Schneider, P., Hofmann, K., Mattmann, C. & Tschopp, J. Interaction of Fas(Apo-1/CD95) with proteins implicated in the ubiquitination pathway. *FEBS Lett.* **412**, 102–106 (1997).
- Wright, D. A., Futcher, B., Ghosh, P. & Geha, R. S. Association of human Fas (CD95) with a ubiquitin-conjugating enzyme (UBC-FAP). *J. Biol. Chem.* **271**, 31037–31043 (1996).
- Lalio, V. S., Vergarajaregui, S., Pulido, D. & Sandoval, I. V. The insulin-sensitive glucose transporter, GLUT4, interacts physically with Daxx. Two proteins with capacity to bind Ubc9 and conjugated to SUMO-1. *J. Biol. Chem.* **277**, 19783–19791 (2002).
- Shepherd, P. R. & Kahn, B. B. Glucose transporters and insulin action—implications for insulin resistance and diabetes mellitus. *N. Engl. J. Med.* **341**, 248–257 (1999).
- Perera, H. K. I. Syntaxin 6 regulates Glut4 trafficking in 3T3-L1 adipocytes. *Mol. Biol. Cell* **14**, 2946–2958 (2003).
- Katome, T. Use of RNA-interference-mediated gene silencing and adenoviral overexpression to elucidate the roles of AKT/protein kinase B isoforms in insulin actions. *J. Biol. Chem.* **278**, 28312–28323 (2003).
- Inoue, M., Chang, L., Hwang, J., Chiang, S. H. & Saltiel, A. R. The exocyst complex is required for targeting of Glut4 to the plasma membrane by insulin. *Nature* **422**, 629–633 (2003).
- Bose, A. *et al.* Glucose transporter recycling in response to insulin is facilitated by myosin Myo1c. *Nature* **420**, 821–824 (2002).
- Wilson, V. G. & Rangasamy, D. Intracellular targeting of proteins by sumoylation. *Exp. Cell Res.* **271**, 57–65 (2001).
- Naviaux, R. K., Costanzi, E., Haas, M. & Verma, I. M. The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses. *J. Virol.* **70**, 5701–5705 (1996).
- Tordjman, K. M., Leingang, K. A., James, D. E. & Mueckler, M. M. Differential regulation of two distinct glucose transporter species expressed in 3T3-L1 adipocytes: effect of chronic insulin and tolbutamide treatment. *Proc. Natl Acad. Sci. USA* **86**, 7761–7765 (1989).
- Bogan, J. S. & Lodish, H. F. Two compartments for insulin-stimulated exocytosis in 3T3-L1 adipocytes defined by endogenous ACRP30 and GLUT4. *J. Cell Biol.* **146**, 609–620 (1999).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank P. Bickel for help with cDNA library construction, G. Paradis for assistance with flow cytometry, J. Cresswell for technical assistance, and N.-W. Chi, J. Avruch, J. Dziura and K. Calia for advice and encouragement. We thank I. Verma for the pCL-Eco plasmid, G. Nolan for retroviral packaging cell lines, M. Charron for GLUT4 antisera, M. Krieger for CHO cells expressing the ecotropic murine retrovirus receptor, and N. Watson and S. Zarnegar for help with microscopy. This work used the W.M. Keck Foundation Biological Imaging Facility at the Whitehead Institute, and the Center for Cell Imaging at Yale University. This work was supported by a New Investigator Award from the Chestnut Hill Charitable Foundation to J.S.B., by a Career Development Award from the American Diabetes Association to J.S.B., by a Pilot and Feasibility Award from the NIH-sponsored Boston Area Diabetes Endocrinology Research Center to J.S.B., by a postdoctoral fellowship from the Ares-Serono Foundation to T.-S.T., and by NIH grants to H.F.L. and J.S.B.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.B. (jonathan.bogan@yale.edu). The cDNA sequences of the short and long TUG forms have been deposited in GenBank under accession numbers AY349132 and AY349133, respectively.

The tobacco aquaporin NtAQP1 is a membrane CO₂ pore with physiological functions

Norbert Uehlein¹*, Claudio Lovisolo²*, Franka Siefritz¹ & Ralf Kaldenhoff¹

¹Institute of Botany, Schnittspahnstrasse 3, D-64287 Darmstadt, Germany

²Department of Arboriculture and Pomology, University of Turin, I-10095 Grugliasco, Italy

* These authors contributed equally to this work

Aquaporins, found in virtually all living organisms, are membrane-intrinsic proteins that form water-permeable complexes. The mammalian aquaporin AQP1 has also shown CO₂ permeability when expressed heterologously in *Xenopus* oocytes¹, although whether this is a biochemical curiosity or of physiological significance is a matter of debate^{2,3}. Here we report that, in the same expression system, a CO₂ permeability comparable to that of the human AQP1 is observed for the tobacco plasma membrane aquaporin NtAQP1. NtAQP1 facilitates CO₂ membrane transport in the homologous plant system at the cellular level, and has a significant function in photosynthesis and in stomatal opening. NtAQP1 overexpression heightens membrane permeability for CO₂ and water, and increases leaf growth. The results indicate that NtAQP1-related CO₂ permeability is of physiological importance under conditions where the CO₂ gradient across a membrane is small, as is the case between the atmosphere and the inside of a plant cell.

CO₂ gas exchange is of pivotal significance, not only in animals but also in plants. CO₂ is the substrate for photosynthetic carbon fixation and in its reduced form, as a sugar compound, is the basis for life. However, the key CO₂-reducing enzyme Rubisco has a relatively low affinity for this substrate, and the CO₂ gradient between the atmosphere and the plant is ~400 times less than that between animal lungs and the atmosphere. Nevertheless, the CO₂ exchange rates in mammalian lungs and plants are similar⁴. Consequently, a low resistance to CO₂ diffusion in plants can be expected and a slight variation of this resistance—for example, by a change of membrane CO₂ transport rate—should immediately affect reactions that are limited by CO₂ availability, such as the photosynthesis rate. Owing to the high sequence conservation of aquaporins, it is possible that a plant aquaporin with a CO₂

permeability comparable to the human AQP1 could exist, and facilitation of CO₂ membrane transport could explain the inconsistency between the small CO₂ concentration gradients and CO₂ exchange rates in plants. In fact, the plant aquaporin NtAQP1 does share a high sequence similarity with AQP1 and functionally important amino-acid residues in the pore region are identical⁵. Whether the CO₂ permeability of the human AQP1 is just a biophysical curiosity or has physiological relevance was analysed by comparing transgenic AQP1-null mice with controls². It was found that CO₂ blow-off by the lung was independent of AQP1 expression, and the CO₂ permeability of AQP1 thus appeared irrelevant for CO₂ permeation in mice³. Because the existence of unstirred layers or perfusion-limited conditions might have masked the contribution of AQP1 to CO₂ permeability, concerns were raised against this interpretation⁶.

Using the technique of Nakhoul *et al.*¹, we show here that the plant aquaporin NtAQP1, when expressed in *Xenopus* oocytes, has CO₂ permeability features similar to those of the human AQP1. Furthermore, in contrast to findings in animal systems, the physiological significance of aquaporin CO₂ permeability can be attributed to NtAQP1 from the data provided in this report.

For the determination of CO₂ membrane transport rates, *Xenopus* oocytes were injected with a solution of carbonic anhydrase, an enzyme that accelerates the conversion of CO₂ to HCO₃⁻. In these oocytes, CO₂ membrane transport rather than the conversion reaction is rate limiting for HCO₃⁻ accumulation¹. As a result, CO₂ transport into the cells caused a decrease in intracellular pH (pH_i) in controls injected with carbonic anhydrase only, and also in NtAQP1-expressing/carbonic anhydrase oocytes (Fig. 1). From the rates of pH_i decrease, we found that the CO₂ uptake rates of controls and the initial CO₂ uptake rates of oocytes overexpressing NtAQP1 were, respectively, 45% higher (*n* = 10) and in a range comparable to those of the human AQP1. Therefore, in addition to its permeability to water and glycerol⁷, NtAQP1 is also a CO₂ transporter in the oocyte expression system. Results from heterologous expression can be transferred to the homologous system only with many restrictions. For analysis of NtAQP1 function on CO₂ transport in the homologous system, NtAQP1 antisense tobacco lines with a severe impairment in NtAQP1 expression were used⁸, as well as plants with an NtAQP1 coding region under the control of a tetracycline-inducible promoter⁹ (TET-NtAQP1). Thus, plants with a relatively low (antisense NtAQP1 lines) or high (TET-

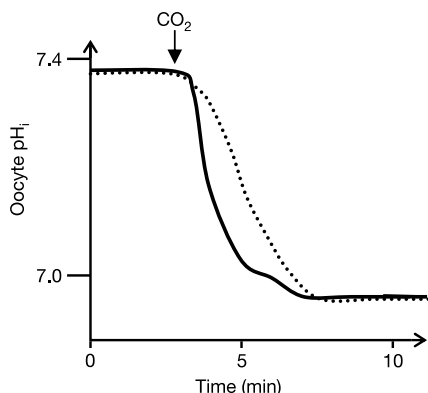


Figure 1 Representative kinetics of the pH decrease in a *Xenopus* oocyte injected with carbonic anhydrase only (dotted line) or with additional expression of NtAQP1. Initial rates for the oocyte overexpressing NtAQP1 indicate a 45% increase in CO₂ plasma membrane permeability.

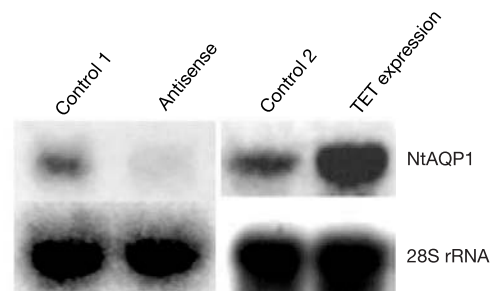


Figure 2 Equal amounts of total RNA (10 µg) were loaded on a gel, blotted and hybridized to a gene-specific probe for NtAQP1 or to 28S-rRNA as loading control. RNA was isolated from controls or NtAQP1-transformed lines: *Nicotiana tabacum* var. *samsun* transformed with a 35S-GUS construct for comparison with antisense lines constructed in the same variety (Control 1) and transformed with a 35S-NtAQP1 antisense construct (Antisense)¹¹; or *N. tabacum*, Hö 20.20 transformed with a 35S-tetracycline-repressor construct and a construct containing an NtAQP1 gene under control of a tetracycline-repressor-regulated promoter. Plants were treated with tetracycline (TET expression) or irrigated with water. After application of tetracycline to Hö 20.20 lines without the NtAQP1 construct (Control 2), expression of NtAQP1 was comparable to that of the watered plants mentioned above.

NtAQP1 lines treated with tetracycline) NtAQP1 expression were available as indicated by the northern blots in Fig. 2. Cellular CO_2 transport was monitored through the incorporation of $^{14}\text{CO}_2$ into photosynthetic products of leaf discs. With increasing CO_2 concentrations, differences in ^{14}C incorporation rates were obtained in all tobacco lines. However, ^{14}C incorporation was dependent on the NtAQP1 expression: a low NtAQP1 level resulted in lower incorporation and an increased NtAQP1 level resulted in higher incorporation rates compared with controls. Differences in absolute uptake values of controls in Fig. 3a, b result from differences in the photosynthetic capacity of the lines used: var. *samsun* (collection of K.W. Mundry, Stuttgart, Germany) for the antisense lines; line Hö 20.20 (ref. 10) for TET-NtAQP1. The evidence suggests that NtAQP1 also acts as a CO_2 membrane-transport-facilitating protein in the homologous system of tobacco. Whether a variation of cellular CO_2 permeability with NtAQP1 function in intact plants is also relevant for physiological processes was initially investigated on leaf stomatal opening kinetics, because these are known to be highly sensitive to changes in CO_2 concentrations. Guard cells and the cells in their vicinity seem to receive a CO_2 concentration signal¹¹; elevated CO_2 typically stimulates stomatal closure, whereas lowered CO_2 concentrations have the opposite effect¹². CO_2 concentration differences change apoplastic pH and the activity of plasma membrane ion channels and H^+ -ATPase¹³. The mechanism of stomatal closure as a result of increased CO_2 concentration has been explained by the malate hypothesis¹⁴, which involves a release of malate and anions from mesophyll and guard cells together with cations. An accumulation of NtAQP1 was obtained in these tissues¹⁵, and we speculated that the NtAQP1 CO_2 permeability function would be related to the signal that determines stomatal guard cell movement. NtAQP1 expression should thus affect the CO_2 concentration that is sensed, and the plant's response should become easily measurable in the very sensitive CO_2 -concentration-

regulated stomatal system. The respective experiments were conducted with leaves on intact plants as well as with detached leaves; this allowed the exclusion of the effects of reduced root hydraulic conductivity caused by the lack of NtAQP1 in roots⁸. No significant differences between attached and detached leaves were observed, indicating that the results depended on a leaf-located mechanism and not on effects in the roots. When NtAQP1 antisense and control plants were transferred from darkness to light under ambient CO_2 concentrations (380 p.p.m.), light-induced stomatal opening occurred with approximately the same velocity (Fig. 4a). This suggests that the mechanism determining membrane water flux rates, in this case those of guard cell membranes, was not affected. This argument was supported by the kinetics of abscisic-acid-induced stomatal closing, which were very similar in controls and in plants with aberrant NtAQP1 expression (data not shown). The opening kinetics of stomata in both plant lines display a characteristic overshoot, which has been interpreted by Raschke¹⁶ as a delayed negative regulation after the positive (light) opening signal. The subsequent oscillation in antisense plants resembled stomatal behaviour under artificial and abrupt CO_2 concentration changes. It indicates a disproportion between positive (light) and negative (CO_2) signals, and it is characteristic of a persistent delay in the negative feedback mechanism. Increasing the cellular CO_2 permeability by overexpression of NtAQP1 notably decreased the overshoot reaction (Fig. 4b). During the light period, the CO_2 concentration in the leaves, $[\text{CO}_2]_i$, was similar in all plant lines (Table 1); however, net photosynthesis was reduced in antisense lines (57%) and increased in overexpressing lines (136%). This indicates that the reduced CO_2 availability in antisense plants was rate limiting for photosynthesis, confirming detailed physiological studies⁴. Consequently, an increase in CO_2 membrane permeability raised the average net photosynthesis. Under increased CO_2 conditions (820 p.p.m.), when the CO_2 concentration gradient between

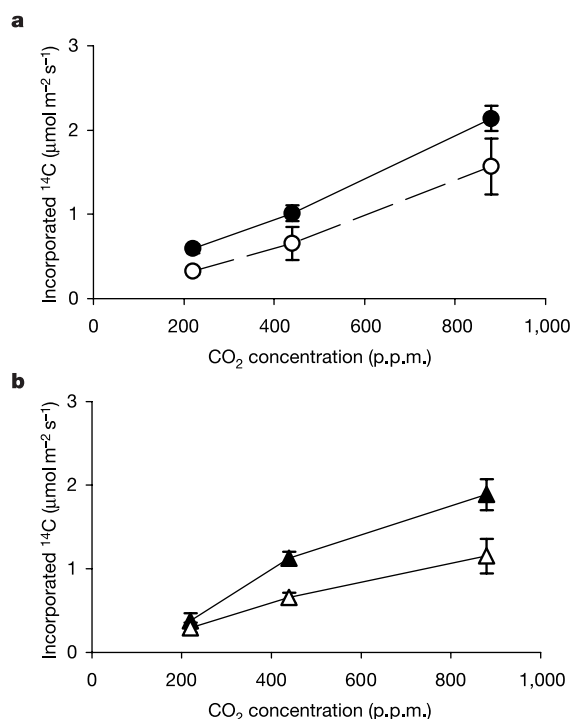


Figure 3 Incorporation of ^{14}C into acid-precipitated compounds after incubation with $^{14}\text{CO}_2$. Continuous line/filled circles, controls as described in Fig. 2. **a**, Results for antisense lines (long dashes/open circles). **b**, Results for tetracycline-induced NtAQP1 overexpression (dotted line/filled triangles).

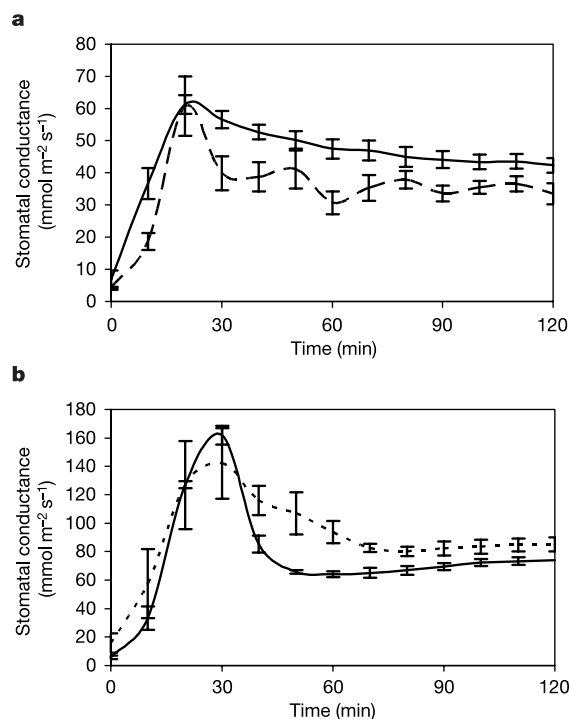


Figure 4 Kinetics of stomatal conductance from plants kept in the darkness and illuminated with $250 \mu\text{mol m}^{-2} \text{s}^{-1}$ PPFD at time 0. Controls (continuous line) as described in Fig. 2. **a**, Results for antisense lines (long dashes); **b**, results for tetracycline-induced NtAQP1 overexpression (dotted line).

Table 1 Intercellular CO₂ concentration, net photosynthesis and stomatal conductance of plants with altered NtAQP1 expression

	Ambient CO ₂ (380 p.p.m.)			High CO ₂ (820 p.p.m.)		
	[CO ₂] _i (p.p.m.)	Net photosynthesis (μmol m ⁻² s ⁻¹)	Stomatal conductance (μmol m ⁻² s ⁻¹)	[CO ₂] _i (p.p.m.)	Net photosynthesis (μmol m ⁻² s ⁻¹)	Stomatal conductance (μmol m ⁻² s ⁻¹)
Control 1	248 ± 18	3.35 ± 0.8	44.1 ± 2.6	579 ± 91	2.43 ± 1.4	17.2 ± 1.3
Antisense	253 ± 47	1.93 ± 0.4	33.6 ± 2.4	536 ± 97	2.42 ± 0.6	14.7 ± 1.4
Control 2	251 ± 18	5.17 ± 0.7	69.4 ± 2.4	508 ± 52	4.67 ± 1.0	26.0 ± 1.2
TET expression	248 ± 13	7.04 ± 0.7	82.3 ± 4.8	459 ± 92	8.48 ± 1.2	35.4 ± 3.2

Calculated CO₂ concentration in the stomatal cavity, [CO₂]_i, and measurements of net photosynthesis and stomatal conductance (illumination for 90 min). Data were obtained from three different transformed lines, two plants each ($n = 6$, $\pm$ s.d.). 'Control 1', 'Antisense', 'Control 2' and 'TET expression' refer to plant lines described in Fig. 2.

the atmosphere and the inside of the cell is steeper, net photosynthesis in antisense NtAQP1 plants reached a level comparable to that of controls. Apparently, an increase in membrane resistance to CO₂ transport was not relevant under these conditions. In the NtAQP1-overexpressing plants, the opposite result was obtained: net photosynthesis was still increased (181%) or was even higher than that at ambient CO₂ concentrations, indicating that a decrease in membrane resistance to CO₂ transport promoted photosynthetic rates by raising the apparent CO₂ availability in the cells. Compared with other plant species, the absolute photosynthesis rates of the tobacco plants seem relatively low. However, the relative differences in photosynthesis rates between the plants with different NtAQP1 levels, and within a certain range of CO₂ concentrations, indicates that plant aquaporin functions as a CO₂ facilitator. Further data are necessary to calculate absolute values for the contribution of aquaporins to CO₂ permeability of the plasma membrane of mesophyll cells.

Eventually, the overexpression of NtAQP1 resulted in improved leaf growth of greenhouse-grown tobacco. Leaves of TET-NtAQP1 plants treated for a period of 6 days with tetracycline had a longitudinal leaf growth velocity of 1.7 ± 0.3 cm per day. Comparable leaves of TET-NtAQP1 plants without tetracycline treatment or 'Control 2' plants (see Fig. 2) increased by only 0.6 ± 0.1 and 0.7 ± 0.1 cm per day, respectively. By contrast, neither plant height nor root mass showed significant growth differences. The increased growth velocity of leaves from NtAQP1-overexpressing tobacco could be a result of the higher photosynthetic capacity, but could also be a synergistic effect of the changed CO₂ and water permeability in the whole plant. Clearly, NtAQP1 overexpression supported plant growth under optimal greenhouse conditions.

The results of this study provide evidence that the plant aquaporin NtAQP1 acts as a CO₂ transporter, not only in a heterologous *Xenopus* expression system but also in the plant itself. Expression of this protein had consequences for CO₂-dependent cellular and physiological processes. Our work provides evidence that under the environmental conditions of a plant, which are characterized by a low CO₂ concentration gradient, the CO₂ permeability of aquaporins is not merely a biochemical curiosity but is of substantial physiological significance. In other organisms, a comparable physiological function of aquaporins as membrane CO₂ facilitators might be found where similar low CO₂ gradients exist. □

Methods

Oocyte expression and CO₂ permeability measurements

Stage V to VI oocytes of *Xenopus laevis* were injected with 50 ng of antisense RNA or sterile water as a control⁷ and kept in an incubation chamber at 16 °C. Aquaporin function was observable after 3–5 days. The CO₂ permeability of the oocyte plasma membrane was measured as described¹, except that the CO₂-containing bath solution was aerated with 5% CO₂/95% O₂. pH micro-electrodes had sensitivities of 58.00 ± 0.58 mV per pH unit. Values calculated from the initial pH_i decrease of control oocytes ($n = 10$) were set to 1 and the values of NtAQP1-expressing oocytes ($n = 10$) were related to these.

Plant growth

Plants were cultivated under greenhouse conditions in shaded light (200–400 μmol m⁻² s⁻¹).

Tetracycline-inducible NtAQP1-overexpressing plants

The NtAQP1 coding region (AJ001416) was cloned into the pBinHyg-Tx vector¹⁶. This vector, a pBin19 derivative, contains a modified 35S-CaMV promoter for effective overexpression of the inserted gene, a binding site for the tetracycline-sensitive repressor (TetR) and a *hptII* gene. pBinHyg-Tx was kindly provided by C. Gatz. Competent agrobacteria LBA4404 were transformed by heat shock. TetR-tobacco leaf discs were transformed following the protocol of ref. 17. The T2 generation was tested for tetracycline-induced transcription before conducting physiological experiments. Plants were grown under normal greenhouse conditions.

Tetracycline treatment

For the induction of NtAQP1 expression in the whole plant, soil-grown plants of ~8 weeks old were watered with 10 mg l⁻¹ tetracycline once a day. Experiments were performed after 4 days of tetracycline treatment. For leaf growth measurements, leaves 10 cm in length were used and growth was monitored for 6 days. Growth velocity was constant until the leaf reached a final length of ~20 cm; it then decreased to almost zero.

¹⁴C uptake in tobacco leaf discs

¹⁴C uptake was determined following a protocol provided by ref. 18. Six leaf discs without epidermis (diameter, 6 mm) were prepared from ~2-week-old leaves of 10–12 cm length and incubated in a 10 ml glass centrifuge tube in 0.9 ml incubation buffer (50 mM HEPES pH 7.6, 1 mM MgCl₂, 350 mM sorbitol, 1 mM dithiothreitol, 1% bovine serum albumin), vacuum infiltrated and pre-illuminated in a plexiglas shaking rack. After 5 min, ¹⁴C-KHCO₃ (1 μCi μmol⁻¹) was added and illumination was continued for another 3 min. The reaction was stopped by adding 0.5 ml of glacial acetic acid, and the leaf discs were homogenized using sea sand and a glass rod; 0.5 ml of the supernatant was transferred to a scintillation vial and dried. Again 0.5 ml of glacial acetic acid was added, and after a final drying cycle the samples were dissolved in 0.5 ml of double-distilled water and mixed with 4 ml of scintillation cocktail (Emulsifier-Safe, Packard Bioscience). Radioactivity was measured by scintillation counting (TRI-CARB Liquid Scintillation Analyzer 1900 CA, Packard Instruments). The results presented here are the average of six repetitions $\pm$ s.d. for each given CO₂ concentration.

Gas exchange measurements

Gas exchange measurements and CO₂ feeding experiments in tobacco plants were performed using HCM-1000 open-flow gas circuit photosynthesis systems in combination with CO₂ dosing units (Walz). H₂O and CO₂ concentrations at the inlet and outlet of the cuvette were measured using differential infrared gas analysers (BINOS-100/4PS, Walz); the gas flow was adjusted to 600 ml min⁻¹. The cuvette area was 2.5 cm² and white light of 250 μmol m⁻² s⁻¹ photosynthetic photon flux density (PPFD) was used. Leaf temperature was kept at 25.5 ± 0.9 °C throughout. Data were recorded at 1 min intervals using DA-1000 software (Walz). Data analysis and calculations were carried out using Microsoft Excel (Microsoft Corporation).

Data analysis

Statistically significant differences between plants with altered NtAQP1 expression and the respective controls were analysed using Student's *t*-test ($P < 0.01$). Data are given as means $\pm$ s.d.

Received 21 May; accepted 28 August 2003; doi:10.1038/nature02027.

Published online 28 September 2003.

- Nakhoul, N. L., Davis, B. A., Romero, M. F. & Boron, W. F. Effect of expressing the water channel aquaporin-1 on the CO₂ permeability of *Xenopus* oocytes. *Am. J. Physiol. Cell Physiol.* **43**, C543–C548 (1998).
- Verkman, A. S. Does aquaporin-1 pass gas? An opposing view. *J. Physiol. (Lond.)* **542**, 31 (2002).
- Yang, B., Fukuda, N., van Hoek, A., Matthay, M. A. & Verkman, A. S. Carbon dioxide permeability of aquaporin-1 measured in erythrocytes and lung of aquaporin-1 null mice and in reconstituted proteoliposomes. *J. Biol. Chem.* **275**, 2686–2692 (2000).
- Evans, J. R. & von Caemmerer, S. Carbon dioxide diffusion inside leaves. *Plant Physiol.* **110**, 339–346 (1996).
- de Groot, B. L. & Grubmüller, H. Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF. *Science* **294**, 2353–2357 (2001).
- Cooper, G. J., Zhou, Y., Bouyer, P., Grichchenko, I. I. & Boron, W. F. Transport of volatile solutes through AQP1. *J. Physiol. (Lond.)* **542**, 17–29 (2002).
- Biela, A. et al. The *Nicotiana tabacum* plasma membrane aquaporin NtAQP1 is mercury-insensitive and permeable for glycerol. *Plant J.* **18**, 565–570 (1999).
- Siefritz, F., Tyree, M. T., Lovisolo, C., Schubert, A. & Kaldenhoff, R. PIP1 plasma membrane aquaporins in tobacco: from cellular effects to function in plants. *Plant Cell* **14**, 869–876 (2002).

9. Gatz, C. Chemical control of gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **48**, 89–108 (1997).
10. Gatz, C., Frohberg, C. & Wendenburg, R. Stringent repression and homogeneous de-repression by tetracycline of a modified CaMV 35S promoter in intact transgenic tobacco plants. *Plant J.* **2**, 397–404 (1992).
11. Raschke, K. Stomatal action. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **26**, 309–340 (1975).
12. Mansfield, T. A., Hetherington, A. M. & Atkinson, C. J. Some current aspects of stomatal physiology. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **41**, 55–75 (1990).
13. Hedrich, R. *et al.* Changes in apoplastic pH and membrane potential in leaves in relation to stomatal responses to CO₂, malate, abscisic acid or interruption of water supply. *Planta* **213**, 594–601 (2001).
14. Hedrich, R. *et al.* Malate-sensitive anion channels enable guard-cells to sense changes in the ambient CO₂ concentration. *Plant J.* **6**, 741–748 (1994).
15. Otto, B. & Kaldenhoff, R. Cell-specific expression of the mercury-insensitive plasma-membrane aquaporin NtAQP1 from *Nicotiana tabacum*. *Planta* **211**, 167–172 (2000).
16. Raschke, K. Saturation kinetics of velocity of stomatal closing in response to CO₂. *Plant Physiol.* **49**, 229–234 (1972).
17. Gallois, P. & Marinho, P. Leaf disk transformation using *Agrobacterium tumefaciens*—expression of heterologous genes in tobacco. *Methods Mol. Biol.* **49**, 39–48 (1995).
18. Kaiser, W. M. Correlation between changes in photosynthetic activity and changes in total protoplast volume in leaf tissue from hygro-, meso- and xerophytes under osmotic stress. *Planta* **154**, 538–545 (1982).

Acknowledgements We thank W. M. Kaiser, M. Eckert and A. Schubert for discussion and help in experimental design.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.K. (kaldenhoff@bio.tu-darmstadt.de).

Global analysis of protein expression in yeast

Sina Ghaemmaghami^{1,2}, Won-Ki Huh^{1,3}, Kiowa Bower^{1,2}, Russell W. Howson^{1,3}, Archana Belle^{1,3}, Noah Dephoure^{1,3}, Erin K. O'Shea^{1,3} & Jonathan S. Weissman^{1,2}

¹Howard Hughes Medical Institute, ²Departments of Cellular & Molecular Pharmacology and ³Biochemistry & Biophysics, University of California—San Francisco, San Francisco, California 94143-2240, USA

The availability of complete genomic sequences and technologies that allow comprehensive analysis of global expression profiles of messenger RNA^{1–3} have greatly expanded our ability to monitor the internal state of a cell. Yet biological systems ultimately need to be explained in terms of the activity, regulation and modification of proteins—and the ubiquitous occurrence of post-transcriptional regulation makes mRNA an imperfect proxy for such information. To facilitate global protein analyses, we have created a *Saccharomyces cerevisiae* fusion library where each open reading frame is tagged with a high-affinity epitope and expressed from its natural chromosomal location. Through immunodetection of the common tag, we obtain a census of proteins expressed during log-phase growth and measurements of their absolute levels. We find that about 80% of the proteome is expressed during normal growth conditions, and, using additional sequence information, we systematically identify mis-annotated genes. The abundance of proteins ranges from fewer than 50 to more than 10⁶ molecules per cell. Many of these molecules, including essential proteins and most transcription factors, are present at levels that are not readily detectable by other proteomic techniques nor predictable by mRNA levels or codon bias measurements.

The diverse chemical nature of proteins makes the development of globally applicable proteomic assays very challenging. We have overcome this obstacle in the yeast *S. cerevisiae* by individually tagging each of its annotated open reading frames (ORFs) with a

high-affinity epitope tag so that the resulting fusion proteins are expressed under the control of their natural promoters. The fusion library allows the immunodetection and immunopurification of the entire yeast proteome using a single antibody, enabling the development of a range of high-throughput functional assays. To allow for the facile construction of epitope-tagged yeast fusion libraries, we synthesized 6,234 pairs of ORF-specific oligonucleotide primers. Each of the oligonucleotide pairs have shared 3' ends that allow for polymerase chain reaction (PCR) amplification of a common insertion cassette, as well as gene-specific 5' ends that allow for the precise introduction, through homologous recombination, of the amplified insertion cassettes as a perfect in-frame fusion at the carboxy-terminal end of the coding region of each gene⁴ (Fig. 1a). The insertion cassettes contained the coding region

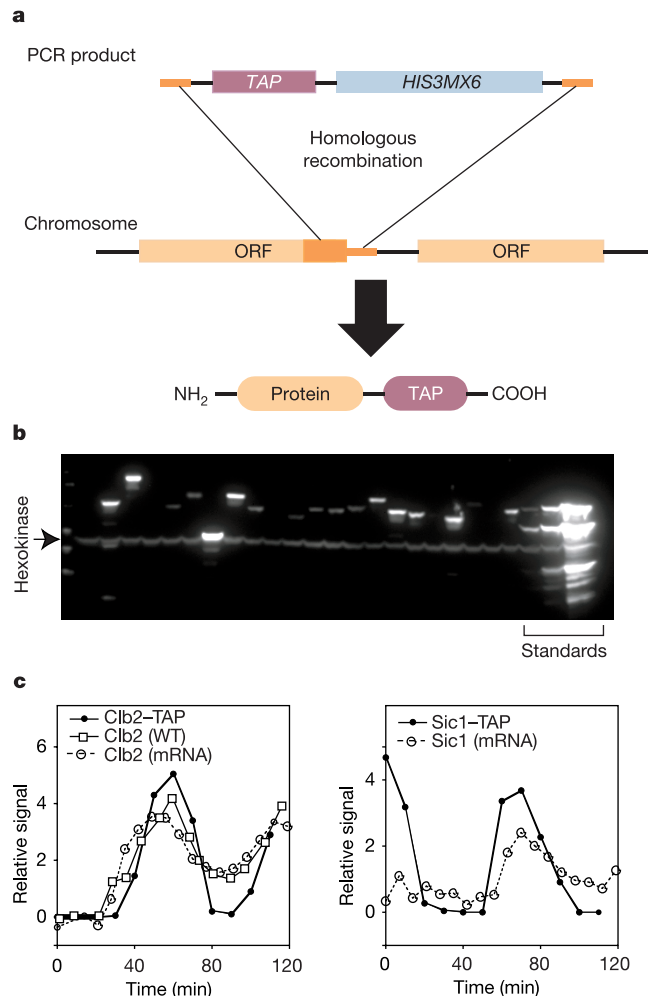


Figure 1 Tagging and detection of the yeast proteome. **a**, Schematic diagram of tagging strategy. **b**, Detection of tagged proteins. Extracts containing TAP-fusion proteins were prepared and analysed by western blots using an anti-CBP antibody (see Supplementary Information). Immunodetection of an endogenous protein (hexokinase) provided a loading control. Serial dilutions of TAP-tagged proteins provided an internal abundance standard (right). **c**, Monitoring dynamic protein levels for two cell-cycle regulated proteins. Strains expressing Clb2– and Sic1–TAP fusions were grown to log-phase and arrested in G1 by α -factor treatment. The cell cycle was resumed by α -factor removal, aliquots were taken at 7-min intervals and levels of the tagged proteins were quantified using western blot analysis (filled circles). For comparison, we include mRNA levels of the two proteins obtained by an earlier microarray analysis²⁹ (open circles) as well as changes in untagged Clb2 protein levels (open squares) obtained using an antibody against the endogenous protein in an untagged strain.

for a modified version of the tandem affinity purification (TAP) tag^{5,6}, which consists of a calmodulin binding peptide, a TEV cleavage site and two IgG binding domains of *Staphylococcus aureus* protein A, as well as a selectable marker (see Supplementary Information). In total, we obtained successful integrants for 98% of all ORFs annotated in the *Saccharomyces* genome database (as of April 2001; <http://www-genome.stanford.edu/Saccharomyces>), including 93% of all essential ORFs⁷ in haploid yeast.

Western blot analysis, using an antibody that specifically recognizes the TAP tag, demonstrated that the large majority (>95%) of detected fusion proteins migrate predominantly as a single band of the approximate expected molecular mass (Fig. 1b). Furthermore, analysis of two known cell-cycle-regulated proteins, Clb2 and Sic1^{8,9}, indicated that the tagging does not hinder their regulated proteolysis by the ubiquitin/proteasome degradation system and that the TAP tag itself is rapidly destroyed during the targeted degradation of the fusion protein (Fig. 1c). These and other data⁶ suggest that the function, regulation and stability of most, but not all (see Supplementary Information), of the proteome is uncompromised by the fused tag.

We observed a protein product for 4,251 of the TAP-tagged ORFs by comprehensive western blot analysis. This set of proteins shows excellent overlap (>90%) with the set of green fluorescent protein (GFP) fusion proteins detected by fluorescence microscopy¹⁰ (Fig. 2a), and together indicate that at least 4,517 proteins are expressed during log-phase growth in rich media. We detect 79% of all essential proteins and 83% of gene products corresponding to ORFs with assigned gene names. By contrast, only 73% of all annotated ORFs expressed a detectable protein product (Fig. 2b). This discrepancy largely results from the presence of spurious ORFs in the annotated yeast genome database stemming from well-known difficulties in distinguishing actual coding regions from fortuitous

short ORFs^{11,12}. For the original annotation of the yeast genome, an arbitrary cut-off of 100 codons was used to qualify ORFs as potential genes, leading to an anomalous peak centred between 100 and 150 amino acids in the sequence length distribution (Fig. 2c, black)¹³ of the genome that is not present in the length distribution of the subset of named genes (Fig. 2c, green). Importantly, although we tagged and analysed all potential ORFs, the length distribution of the subset of observed proteins did not contain the above artefactual peak (Fig. 2c, red), indicating that our analysis of expressed genes has a very low false-positive rate (see also Supplementary Information).

A number of bioinformatics approaches, including recent analyses of the genomic sequences of a number of related yeast species, have been used to distinguish between the real and misannotated ORFs^{14–17}, although the true number and identity of the spurious ORFs remain unclear. Our results offer experimental verification for a large number of hypothetical genes (we observed 1,018 protein products belonging to functionally uncharacterized ORFs), and yields a large, experimentally validated set to evaluate the success of computational methods for identifying falsely annotated genes. By combining a novel metric—termed the codon enrichment correlation (CEC), which evaluates the patterns of codon usage in potential ORFs—with our protein expression data, we identified a

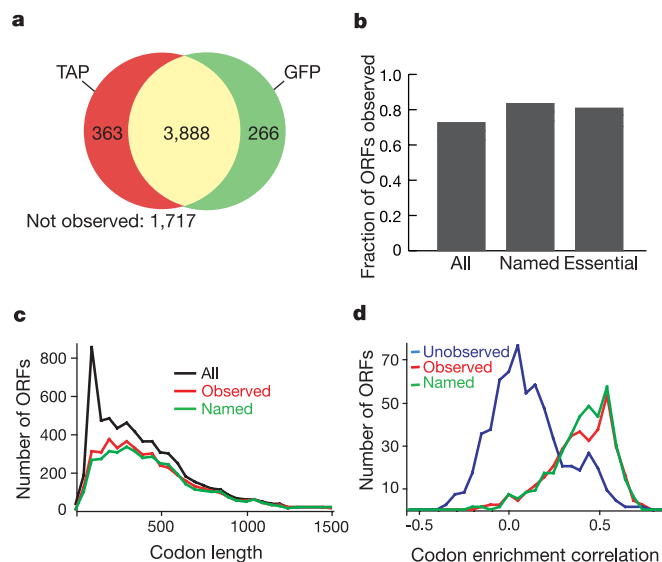


Figure 2 Analysis of proteins expressed during log-phase growth. **a**, Venn diagram comparing sets of proteins detected by western blot of TAP-tagged strains (red), fluorescence microscopy of GFP-tagged strains¹⁰ (green) and both (yellow). **b**, Fraction of the indicated set of ORFs observed in either the TAP-tagged or GFP-tagged libraries. **c**, Size distribution of ORFs, binned by length using 50-codon intervals. The number of ORFs per bin is plotted for the indicated sets of ORFs. **d**, Codon enrichment correlation (CEC) distribution of small ORFs. CECs were calculated for ORFs with lengths from 100 to 150 codons. ORFs were binned according to CEC values using intervals of 0.05 units. The number of ORFs in each bin is plotted for the indicated sets of ORFs. Note, observed proteins have a positive CEC value characteristic of named genes, whereas unobserved ORFs show a major peak centred near a zero value expected for random sequences.

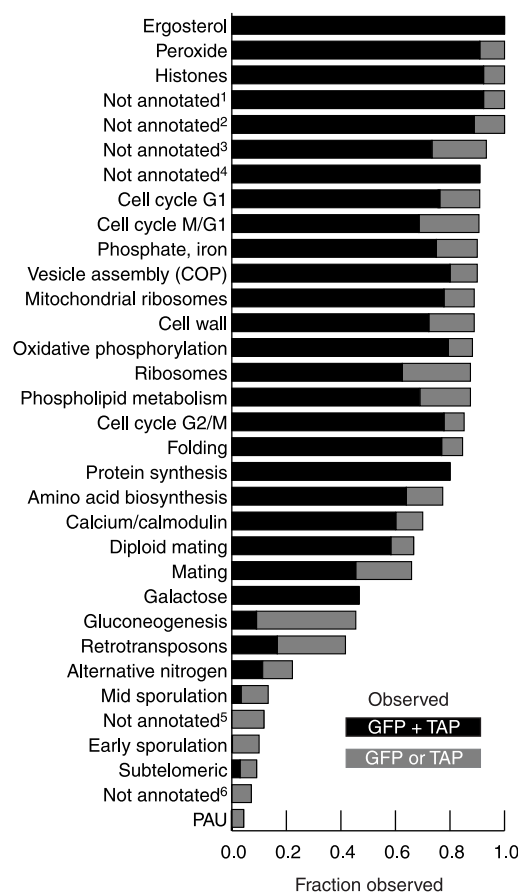


Figure 3 Functional categorization of proteins expressed during log-phase growth in rich medium. 33 modules of co-expressed, functionally related genes were identified by global analysis of ~1,000 microarray data sets^{18,19}. Plotted is the fraction of the ORFs in each module that produced a detectable protein product by TAP western analysis or GFP microscopy¹⁰ alone (grey), or both methods (black). Where possible, modules are annotated by function. The gene composition of the modules can be obtained at <http://barkai-serv.weizmann.ac.il/modules/> using a cut-off threshold of 4.0. 'Not annotated'^{1–6} correspond to modules containing YHR025W, YER103W, YPL016W, YPL180W, YER039C-A and YCL076W, respectively.

set of 525 potentially spurious ORFs (listed in Supplementary Information) that have codon compositions not characteristic of genuine genes and did not yield detectable protein products (Fig. 2d, Methods). On the basis of the CEC distribution of genuine ORFs, we estimate that this list is contaminated by ~20 genuine coding sequences. Our proteomics-based approach complements the comparative genomics strategy for identifying spurious ORFs¹⁶. The large majority (all but seven) of the 496 spurious ORFs suggested by Kellis *et al.*¹⁶ were not observed in our TAP and GFP studies. The set of spurious ORFs that we identified overlaps well with those detected by this cross-species genome study (381 genes were identified as spurious by both studies), and expands the set by 144 ORFs. Among these 144 ORFs are a large number of sequences that overlap with real genes on the opposite strand, and therefore are difficult to distinguish through homology analysis.

After discounting the spurious ORFs, there remain ~1,000 genuine coding regions that did not produce a detectable protein product. To determine if the unobserved proteins belong to classes of genes that are not transcribed during normal log-phase growth conditions, we compared our results with global transcriptional array data. A recent analysis of mRNA expression profiles from ~1,000 published microarray experiments allowed for the identification of 33 'modules' of transcriptionally co-regulated genes^{18,19}. For modules that are expressed in log phase (for example, those coding for housekeeping functions, such as ergosterol and amino-acid biosynthesis and cell cycle), we were able to detect the large

majority of the protein products (Fig. 3). By contrast, modules composed of genes involved in functions required only under specialized conditions (for example, meiosis/sporulation and alternative nitrogen utilization) generally produced few detectable proteins.

We took advantage of the fact that all gene products were detected using the same epitope/antibody interaction to measure the absolute abundance of each of the tagged proteins using quantitative western blot analyses. This effort was facilitated by the inclusion of internal standards in each gel (Fig. 1b). We find that the levels of different proteins show an enormous dynamic range, varying from fewer than 50 to more than 10^6 molecules per cell (Fig. 4a, b). The results show that previous efforts to quantify protein levels using two-dimensional gel electrophoresis or mass spectrometry were strongly biased towards the detection of abundant proteins (Fig. 4a, see also Supplementary Fig. S3)^{20–23}. For example, a recent study using mass spectrometry and isotope labelling succeeded in quantitatively monitoring changes in the abundance of 688 yeast proteins²². For the most abundant proteins (>50,000 molecules per cell) the coverage was excellent (~60%), whereas for the 75% of the proteome that is present at fewer than 5,000 molecules per cell, only 8% of the proteins were observed. Another mass-spectrometry effort that focused on detecting, without directly quantifying, the complement of proteins in log-phase yeast²³ observed a larger number (1,484) of proteins, although it was also biased towards abundant proteins (90% of the proteome present at >50,000

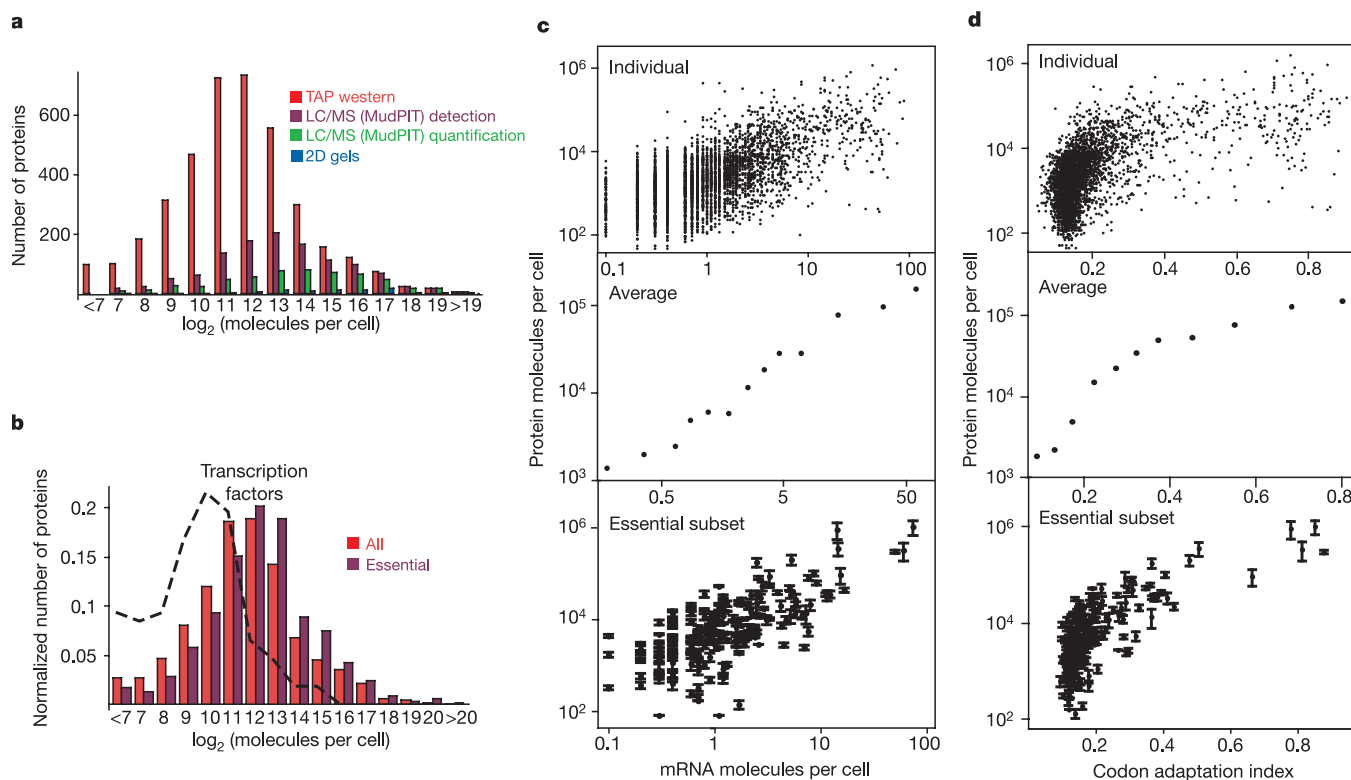


Figure 4 Abundance distribution of the yeast proteome. **a**, Distribution of yeast proteins observed by TAP/western-blot (red), liquid chromatography/mass spectrometry multidimensional protein identification technology (LC/MS MudPIT) analysis focusing on comprehensive detection²³ (purple) and quantitative analysis²² (green), and combined results from 2 two-dimensional (2D) gel analyses^{20,21} (blue). The bins are $\log_2$ increments with upper boundaries indicated. **b**, Normalized abundance distribution of observed proteins (red), essential proteins (purple) and transcription factors (dashed line). **c**, The relationship between steady-state mRNA and protein levels. Top plot, abundance of each protein is plotted against its mRNA level determined by microarray analysis²⁵. Middle plot,

ORFs are sorted according to mRNA levels, and binned into successive groups with cut-offs of 0.25, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0, 4.0, 5.0, 10, 20, 50 and 100 molecules per cell. For each bin, the mean protein abundance is plotted against the mean mRNA level. Bottom plot, protein versus mRNA relationship for a subset of essential soluble proteins (see Supplementary Information). Errors represent the standard deviation of three measurements. **d**, Relationship between codon adaptation index (CAI) and protein levels. Individual and averaged protein values are plotted against CAI²⁷. In the middle plot, the values are binned using CAI cut-offs of 0.1, 0.15, 0.20, 0.25, 0.30, 0.35, 0.40, 0.50, 0.60, 0.70, 0.80 and 1.0.

molecules per cell was detected, whereas only 19% of the proteome present at fewer than 5,000 molecules per cell was observed). Our validated list of expressed proteins will help evaluate future advances in mass spectrometry approaches²⁴.

Overall, we observe a significant relationship between mRNA levels, as measured by an earlier microarray analysis of log-phase yeast²⁵, and protein levels (Spearman rank correlation coefficient $r_s = 0.57$). Very abundant mRNAs generally encode for abundant proteins, and the average protein per mRNA ratio remains remarkably constant throughout the full range of mRNA abundances (Fig. 4c, middle, and Supplementary Fig. S4). The average protein per mRNA ratio is 4,800 using this measure of mRNA levels, and is 4,200 using an alternative mRNA abundance measurement based on a microarray analysis comparing mRNA to genomic DNA levels²⁶ (Supplementary Fig. S4). However, individual genes with equivalent mRNA levels can result in large differences in protein abundances (Fig. 4c, top). To assess if this variability was primarily caused by protein measurement error and/or disruption of protein function by the TAP tag, we performed further triplicate measurements of protein abundances on a subset of 206 essential, soluble proteins (See Supplementary Information); the selected strains grew robustly, showing that the tagged proteins were functional. This subset also shows a high degree of protein to mRNA variability relative to our measurement error, indicating that the large differences in individual protein to mRNA ratios are not due primarily to noise in the protein abundance measurements or disruption of the protein by the tag (Fig. 4c, bottom). However, the correlation between mRNA and protein levels is somewhat greater ($r_s = 0.66$), suggesting that the disruption of protein by the TAP tag or difficulty in analysing membrane proteins may have contributed to some of the variation. We also observed a significant relationship ($r_s = 0.55$) between protein abundance and codon usage as measured by the codon adaptation index (CAI)²⁷. Protein abundances drop rapidly for genes with CAI values < 0.2 , explaining the difficulty that previous proteomic approaches have typically had in detecting these proteins²². But on an individual gene basis, there is great variability that is also present in the subset of more carefully measured essential, soluble proteins (Fig. 4d).

A number of observations support the argument that the full range of abundances detected in this study, including the very low expression levels, represent functionally significant amounts of the proteins. First, the analysis of transcription modules (Fig. 3) indicates that within groups of genes that are turned off during log-phase growth the corresponding proteins are not observed, even at residual levels. Second, the abundance distribution profile of the entire yeast proteome (Fig. 4b, red) is similar to the profile of the portion of the proteome whose function is required for survival under standard growth conditions (Fig. 4b, purple). This suggests that, in general, functional proteins are not under-represented amongst low-abundance proteins. Third, there are entire classes of functionally important proteins, such as transcription factors (Fig. 4b, line) and cell-cycle proteins (Supplementary Fig. S5), that are present at very low expression levels. Thus the low-abundance proteins detected and quantified in the present study represent a large and functionally important portion of the yeast proteome that is almost entirely invisible to systematic quantitative analysis by other proteomic methods.

The TAP-tagged library now makes it feasible to monitor dynamically the abundance of the yeast proteome through basic cellular events such as the cell cycle and meiosis, and will allow the determination of protein lifetimes. In addition, important subsets of proteins, such as transcription factors, can be readily studied under a more comprehensive set of conditions. This protein-based data will provide critical information for efforts to understand the logic of cellular regulatory circuits, and, by comparison to mRNA levels, the data will give insight into the nature and extent of post-transcriptional regulation. □

Methods

Quantification of protein levels

Cultures (1.7 ml) of tagged strains were grown in 96-well format to log phase, and total cell extracts were examined by SDS-polyacrylamide gel electrophoresis (PAGE)/western blot analysis as described in Supplementary Information. The bands corresponding to the tagged proteins were detected using chemiluminescence and a CCD camera (FluorChem 8800, Alpha Innotech). To control for variation in extraction and loading, each blot was probed with an antibody against endogenous hexokinase in addition to the TAP-specific anti-CBP antibody. Extracts whose hexokinase signals varied by greater than a factor of ~ 2 from the expected value were re-grown and re-analysed. A standard containing a mixture of three TAP-tagged proteins (Pgk1, Cdc19, Rpl1A) were included in each gel at one-, ten- and 100-fold dilutions. Proteins whose chemiluminescence signals were approaching saturation were re-examined by performing the western blot analysis using a tenfold dilution of the extract and/or lower exposure times during detection. Before the quantitative SDS-PAGE/western blot analysis, strains were ordered on the basis of estimates of TAP abundance from a preliminary dot-blot analysis. In order to provide a standard for the conversion of western signals to absolute protein levels, a TAP-tagged protein (*Escherichia coli* initiation factor A, INFA) was overexpressed in *E. coli* and purified to homogeneity. Yeast extracts containing serial dilutions of INFA ranging from 500 attomoles (which was the limit of detection, see Supplementary Fig. S1) to 25 picomoles were run on a gel along with extracts from 25 different yeast TAP-tagged strains representing the full range of observed protein signals (a second TAP-tagged protein (initiation factor B) was also analysed to ensure that the observed TAP signal was not influenced by the fusion protein). Comparison of the signals generated by these 25 proteins to the known standards allowed the creation of a conversion factor between the observed western blot signals and absolute protein levels. Based on the number of cells ($\sim 1 \times 10^7$) used for the SDS-PAGE/western blot analysis, the protein levels were then converted to measurements of protein molecules per cell.

In order to assess the error in our quantification, a set of 33 proteins with a range of abundances were grown in duplicate cultures, separately extracted and analysed on different gels. The replicate signals showed a linear correlation coefficient of $R = 0.94$, with the pairs of proteins having a median variation of a factor of 2.0. This error analysis does not account for potential alterations in the endogenous levels of the proteins caused by the fused tag, which may be particularly disruptive for small proteins (Supplementary Information) or difficulty in analysing some polytopic membrane proteins by SDS-PAGE. For dynamic measurements of protein levels (for example, the cell-cycle dependence of Clb2 and Sic1 levels shown in Fig. 1c or triplicate measurements in Fig. 4c, d) much smaller errors can be obtained by running the samples being compared side-by-side on a single gel. For quantification in the triplicate measurements shown at the bottom of Fig. 4c, d, serial dilutions of extracts containing purified TAP-tagged INFA were run on each gel.

CEC and identification of spurious ORFs

Codon usage in genuine protein-coding regions deviates systematically from randomly generated ORFs, owing to both preferences in amino-acid composition and biases in the usage of synonymous codons²⁸, and the codon enrichment correlation (CEC) provides a measure of this deviation. To calculate CEC values, we first determined the relative prevalence of the 61 amino acids specifying codons in the 3,753 named ORFs (Supplementary Table S1). The codon usage expected in random sequences was then calculated based on the approximate prevalence of 30% T, 30% A, 20% C and 20% G nucleotides in the yeast genomes. The enrichment of each codon for the positive set is given by dividing its prevalence among the named ORFs by its expected prevalence in random sequences (Supplementary Table S1). Codon enrichments were similarly calculated for each test ORF. The CEC is the linear correlation coefficient (r) between the codon enrichments of the test ORF and the positive set (for examples, see Supplementary Fig. S2). ORFs were designated as spurious if they failed to be detected by both the TAP and GFP analyses, and they had CEC values below a cut-off of 0.25, 0.16, 0.07 or 0.06 for ORFs of size 0–150, 151–200, 201–250 and 251–300 codons, respectively. For ORFs > 150 amino acids, these values were chosen so that $< 4.5\%$ of the ORFs falling below these cut-offs that are not detected by the GFP or TAP analyses are genuine coding sequences. The number of genuine coding sequences contaminating our list of spurious ORFs was estimated for each size range and CEC cut-off by the following equation: $N_{\text{real}} = N_{\text{obs}}R$, where N_{obs} is the number of detected ORFs that have a CEC value below the cut-off, and R is the ratio of unobserved to observed ORFs, as determined by the probability of detecting named ORFs for the given size range.

Received 28 July; accepted 28 August 2003; doi:10.1038/nature02046.

1. Lashkari, D. A. *et al.* Yeast microarrays for genome wide parallel genetic and gene expression analysis. *Proc. Natl Acad. Sci. USA* **94**, 13057–13062 (1997).
2. Collins, F. S., Green, E. D., Guttacher, A. E. & Guyer, M. S. A vision for the future of genomics research. *Nature* **422**, 835–847 (2003).
3. Lockhart, D. J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).
4. Longtine, M. S. *et al.* Additional modules for versatile and economical PCR-based gene deletion and modification in *Saccharomyces cerevisiae*. *Yeast* **14**, 953–961 (1998).
5. Rigaut, G. *et al.* A generic protein purification method for protein complex characterization and proteome exploration. *Nature Biotechnol.* **17**, 1030–1032 (1999).
6. Gavin, A. C. *et al.* Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**, 141–147 (2002).
7. Winzler, E. A. *et al.* Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285**, 901–906 (1999).
8. Schwob, E., Bohm, T., Mendenhall, M. D. & Nasmyth, K. The B-type cyclin kinase inhibitor p40SIC1

- controls the G1 to S transition in *S. cerevisiae*. *Cell* **79**, 233–244 (1994).
9. Grandin, N. & Reed, S. I. Differential function and expression of *Saccharomyces cerevisiae* B-type cyclins in mitosis and meiosis. *Mol. Cell. Biol.* **13**, 2113–2125 (1993).
 10. Huh, W.-K. *et al.* Global analysis of protein localization in budding yeast. *Nature* **425**, 686–691 (2003).
 11. Harrison, P. M., Kumar, A., Lang, N., Snyder, M. & Gerstein, M. A question of size: The eukaryotic proteome and the problems in defining it. *Nucleic Acids Res.* **30**, 1083–1090 (2002).
 12. Goffeau, A. Four years of post-genomic life with 6,000 yeast genes. *FEBS Lett.* **480**, 37–41 (2000).
 13. Das, S. *et al.* Biology's new Rosetta stone. *Nature* **385**, 29–30 (1997).
 14. Kowalczyk, M., Mackiewicz, P., Gierlik, A., Dudek, M. R. & Cebrat, S. Total number of coding open reading frames in the yeast genome. *Yeast* **15**, 1031–1034 (1999).
 15. Zhang, C. T. & Wang, J. Recognition of protein coding genes in the yeast genome at better than 95% accuracy based on the Z curve. *Nucleic Acids Res.* **28**, 2804–2814 (2000).
 16. Kellis, M., Patterson, N., Endrizzi, M., Birren, B. & Lander, E. S. Sequencing and comparison of yeast species to identify genes and regulatory elements. *Nature* **423**, 241–254 (2003).
 17. Cliften, P. *et al.* Finding functional features in *Saccharomyces* genomes by phylogenetic footprinting. *Science* **301**, 71–76 (2003).
 18. Ihmels, J. *et al.* Revealing modular organization in the yeast transcriptional network. *Nature Genet.* **31**, 370–377 (2002).
 19. Bergmann, S., Ihmels, J. & Barkai, N. Iterative signature algorithm for the analysis of large-scale gene expression data. *Phys. Rev. E* **67**, 031902 (2003).
 20. Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Mol. Cell. Biol.* **19**, 1720–1730 (1999).
 21. Futcher, B., Latter, G. I., Monardo, P., McLaughlin, C. S. & Garrels, J. I. A sampling of the yeast proteome. *Mol. Cell. Biol.* **19**, 7357–7368 (1999).
 22. Washburn, M. P. *et al.* Protein pathway and complex clustering of correlated mRNA and protein expression analyses in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **100**, 3107–3112 (2003).
 23. Washburn, M. P., Wolters, D. & Yates, J. R. III Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
 24. Aebersold, R. & Mann, M. Mass spectrometry-based proteomics. *Nature* **422**, 198–207 (2003).
 25. Holstege, F. C. *et al.* Dissecting the regulatory circuitry of a eukaryotic genome. *Cell* **95**, 717–728 (1998).
 26. Wang, Y. *et al.* Precision and functional specificity in mRNA decay. *Proc. Natl Acad. Sci. USA* **99**, 5860–5865 (2002).
 27. Sharp, P. M. & Li, W. H. The codon Adaptation Index—a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Res.* **15**, 1281–1295 (1987).
 28. Grantham, R., Gautier, C. & Gouy, M. Codon frequencies in 119 individual genes confirm consistent choices of degenerate bases according to genome type. *Nucleic Acids Res.* **8**, 1893–1912 (1980).
 29. Spellman, P. T. *et al.* Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273–3297 (1998).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank A. Carroll and F. Sanchez for technical assistance; J. Falvo, L. Gerke, J. Newman and members of the Weissman and O'Shea laboratories for discussions; and N. Barkai for providing data before publication. This work was supported by the Howard Hughes Medical Institute and the David and Lucile Packard Foundation. S.G. is a recipient of the Ruth L. Kirschstein National Research Service Award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.S.W. (jsw1@itsa.ucsf.edu).

corrigendum

Invariant scaling relations across tree-dominated communities

Brian J. Enquist & Karl J. Niklas

Nature **410**, 655–660 (2001).

Equation (1) of this Article was incorrect as printed. The total biomass, M_{Tot} , per unit area is the summation, or integral, across the size distribution of the number of individuals per unit area, multiplied by their body mass. Thus $M_{\text{Tot}} = \int_a^b MN(M)dM$. Because the number of individuals in a given area is an allometric function of their size, M , we can substitute the observed relationship $N = C_m M^{-3/4}$ to yield the community biomass equation:

$$M_{\text{Tot}} = \int_a^b C_m M^{1/4} = \frac{4}{5} C_m (M_a^{5/4} - M_b^{5/4}) \quad (1)$$

This change does not affect any of the reported conclusions or empirical patterns. □

Tails of discovery

A mouse model, high-throughput assays and more for drug research.



Family commitment: BD measures TF with ELISA.

TransFactor Family Kits

BD Biosciences Clontech

www.bdbiosciences.com; www.clontech.com

ELISA-based assays for TF binding studies

BD TransFactor Family Kits are designed for studying DNA-binding activities of structurally and functionally related transcription factors (TFs). These activities can be analysed in multiple samples simultaneously. The kits allow ELISA-based determination of changes in DNA binding of TFs in tissue or whole-cell extracts and nuclear extracts. The families included in the kits regulate important signal transduction pathways that play key roles in physiological and pathological processes. Kits include the AP-2, HIF-1, NF- κ B, ER and PPAR families of transcription factors.

jetSI

Qbiogene

www.qbiogene.com

Transfection reagents for siRNA delivery

jetSI transfection reagents are optimized for efficient delivery of siRNA duplexes to mammalian cells. They provide up to 98% silencing of a transgene at 10 nM siRNA and up to 80% endogenous lamin A/C silencing in a variety of cell lines. Fluorescent derivatives of jetSI for use in tracking intercellular delivery and trafficking of siRNA are also available. jetSi compacts siRNA duplexes into positively charged particles that interact with anionic proteoglycans at the cell surface and enter cells by endocytosis. They protect siRNA duplexes from degradation and promote rapid endosomal release into the cytoplasm. The reagents work effectively both with and without serum.

Monoclonal mouse c-Kit antibody

Zymed

www.zymed.com

Marker for gastrointestinal stromal tumours

The c-Kit proto-oncogene encodes a 145–165 kDa transmembrane receptor (CD117) with tyrosine kinase activity that is structurally related to the platelet-derived growth factor. In normal tissues, the c-Kit protein is expressed

in mast cells and interstitial Cajal cells. c-Kit antibody has been widely used in the identification and diagnosis of gastrointestinal stromal tumours (GISTs). It can also be used with the antibody panel including CD34, S-100, desmin and smooth muscle actin to differentiate between GISTs, true smooth muscle tumours and neural tumours. The antibody has been characterized for immunohistochemical staining of frozen and formalin-fixed, paraffin-embedded tissue sections.

SHIP2

Echelon Biosciences

www.echelon-inc.com

5' PtdIns phosphatase assay kit

Echelon Biosciences has developed a fluorescence polarization-based assay for determining activity of 5' PtdIns phosphates, including type II SH2-domain-containing inositol 5-phosphatase (SHIP2). SHIP2 is a negative regulator of insulin signalling and is emerging as a novel therapeutic target in type 2 diabetes. The assay is based on the interaction of specific phosphoinositide-binding proteins with fluorophore-labelled phosphoinositide and inositol phosphate tracers. The assay is suitable for high-throughput screening applications, and will facilitate screening for novel inhibitors of phosphatases in drug development.

MediaClave

Integra Biosciences

www.integra-biosciences.com

New unit creates a stir

The MediaClave nutrient media preparator and sterilizer gently heats nutrient media in the stainless steel vessel until the sterilization temperature is reached. The stirring mechanism produces homogenous mixtures at speeds of 75 or 150 r.p.m. and the media are then rapidly cooled for dispensing by a water-cooled jacket coupled with a support pressure

compressor. This prevents delayed boiling during the cool-down phase. The MediaClave comes with a 32-bit microprocessor and a menu-controlled graphic display for entering customized programs and displaying sterilization and error reports. An integrated printer documents sterilization and can be connected to a PC via an RS-232 interface. The unit can handle complicated media such as blood and chocolate agar, and the manufacturer claims it is 95% faster than self-cooling models.

MultiScreen-MIC plates

Millipore

www.millipore.com

Optimal migration, invasion and chemotaxis

Millipore's 96-well MultiScreen-MIC filter plates come with 3 μ m, 5 μ m or 8 μ m polycarbonate membranes for use with a variety of cell types, and support both adherent and suspension cell lines. They are optimized for functional cell-based assays including migration, invasion and chemotaxis. The 96-well design offers the same membrane surface as a 24-well plate but with smaller well volume using less reagent. Applications include cancer research where effects of drug leads on migration and invasion properties of cancer cells are studied, and investigation of cell-cell interactions with endothelial cells in co-culture, angiogenesis and transmigration assays.

SwellGel

Pierce

www.piercebc.com

Albumin removal discs

Pierce's SwellGel discs are designed for high-capacity albumin removal from larger (2–8 ml, single-loading) serum samples. The binding capacity of each disc is >20 mg per ml of human serum albumin. Discs hydrate in less than 2 minutes with the addition of water to form a standard equilibrated chromatographic resin bed of approximately 2 ml.



Beta test: Taconic's ERKO-β mouse.

They have different levels of affinity for species-specific albumins. The discs bind human, bovine, swine and sheep albumin, but not mouse or rat albumin. SwellGel discs are supplied in a dried format, eliminating the risk of leaks, and are suitable for standard chromatography applications.

ERKO-β

Taconic

www.taconic.com

Mouse on target for reproduction research

This mouse model can be used to examine the role of oestrogen receptors in both reproductive and non-reproductive development. The *esr2* gene, which encodes a protein of about 60 kDa, is disrupted in the embryonic stem cells, but *esr1* gene, the α -oestrogen receptor, is not and remains functional. Both sexes of the homozygous ERKO-β mouse develop nor-

mally and live to adulthood, but the females experience infertility and have fewer offspring. They show decreases in ovulation and cellular mass of the oocyte cumulus, an increase in ovarian follicle degeneration and an insignificant response to exogenous hormones. The males have normal fertility but when they age, they suffer from prostatic and urinary bladder hyperplasia. Pubescent and young adult mutated mice are more aggressive than wild mice. The mice can be used in studies of fertility, sexual behaviour, the cardiovascular system and bone homeostasis. The ERKO-β mouse was developed using the BK4 sub-line of the E14tG2a embryonic stem cell and C57BL/6 blastocysts for homologous recombination.

Angiogenesis brochure

Sigma-Aldrich

www.sigma-aldrich.com

A good read for cancer researchers

This 56-page brochure contains more than 700 products for angiogenesis research. Researchers will find 220 products for cancer research, including angiostatin and endostatin, two potent inhibitors of angiogenesis and tumour growth. The brochure also offers angiogenic factors, growth factors, chemokines, and receptors that control vascular development, as well as extracellular matrix components, matrix metalloproteinases and

cathepsins, and adhesion molecules. Antibodies for studying the functional role of these components during angiogenesis, along with auxiliary products such as angiogenesis inhibitors, enzyme substrates and inhibitors, bioactive lipids and plasma proteins are also included. Illustrations depict the key features of angiogenic pathways.

MatriPlates

Anachem

www.anachemlifescience.co.uk

These shape up well

These 384- and 1,536-well plates have a novel well design that optimizes high-throughput assay applications such as fluorescence, chemiluminescence, colorimetric and cell-based methods. They employ a truncated, inverted pyramid shape that enhances assay signal through improved light transfer. The shape also helps to direct pipette tips into the well base, allows liquid to flow smoothly into the bottom without trapping air bubbles and facilitates removal of the entire liquid. The MatriPlate range includes black, white and clear plates, Teflon-impregnated polypropylene plates, low-fluorescence material and chemical-resistant plates.

These notes are compiled in the Nature office from information provided by the manufacturers.

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building

4 Crinan Street

London N1 9XW, UK

Tel +44 (0) 20 7843 4961

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs Sales Director:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Spain/

Portugal/ Belgium:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Switzerland:

Amelie Pequignot (4974)

Natureevents:

Paul Constant (4954)

Production Manager: Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager: Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Haraikatamachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Rinoko Asami

e-mail: rasami@naturejpn.com

naturejobs

It could be worse...

Common job complaints voiced by scientists include the tedium of writing grants and the monotony of pipetting. But those who complain about such tasks may benefit from a little perspective. The worst job in science really stinks, quite literally, according to a recently released survey. And the runners-up aren't much better — they variously involve extreme tedium, the possibility of bodily harm or both.

The magazine *Popular Science* canvassed 1,000 scientists for nominations. The title of the worst job on the list, published this month, is 'flatulent odour judge', a position held in the lab of Michael Levitt, associate chief of staff for research at the Minneapolis Veterans Affairs Medical Center. Close seconds include collecting mosquitoes — and bites — for malaria research, analysing stool samples of dysentery victims and working in Biosafety Level 4 laboratories that handle Ebola and anthrax samples.

But for historians of science — and people who need to put things into perspective — these positions aren't that bad, and are certainly nothing new, the article notes. Pre-med student Stubbs Ffirth (1784–1820) consumed bodily fluids of yellow-fever victims to pinpoint its cause — and failed, as the samples came from late-stage patients who were no longer contagious. And Marie Curie (1867–1934) exposed herself to radiation, suffering crippling pain, anaemia, cataracts and, eventually, fatal leukaemia. She at least received the Nobel prize in 1911 for discovering polonium and radium, which was some consolation.

The willingness of scientists such as Curie, the workers in Levitt's lab and the other 'winners' to put themselves in discomfort or at risk underscores a truth. They do it because it's worth doing. And as odious and noxious as those tasks can be, we can think of worse — like being unemployed. Or working in a profession that you never intended to and don't enjoy.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

Singapore's big push
for biology

p746

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

ANNOUNCEMENTS

EVENTS

Filling Biopolis Singapore

Covering 18.5 hectares and encompassing seven buildings, Biopolis is a testament to the Singapore government's commitment to biological research, its proposed 'fourth pillar' of the economy. Officially opening next week, this massive research facility represents one of the largest international scientific recruitment drives ever undertaken in Southeast Asia. Built at a cost of some S\$500 million (US\$290 million), the complex will eventually play host to more than 1,500 scientists.

But this is just the beginning. Singapore's drive to attract high-calibre researchers from around the world will see it expand its scientific infrastructure over the next few years. Biopolis is just part of a larger development covering 200 hectares that will also hold Fusionopolis, a media and information-technology tower, as well as affordable housing for scientists working nearby.

Such levels of investment emphasize the government's commitment to science — but is it enough to entice Westerners away from their labs? In its favour, the city-state's official language is English. However, as well as the long distance from 'home', one other factor that may dissuade Western scientists from heading east is the autocratic reputation of Singapore's government. The centralized nature of the administration is a double-edged sword — on the one hand it allows rapid action, such as the construction of Biopolis in under two years, but on the other it has the potential to stifle scientific autonomy.

So far, Biopolis is seeing an influx of big names to head its research institutes, as well as the beginnings of corporate interest in the complex. Novartis, which plans to move its Institute for Tropical Diseases into one of two Biopolis buildings designated for private



Growth industry: Singapore is investing heavily in science.

tenants, is fulfilling both counts. The institute's director is Alex Matter, former head of oncology research at Novartis, who spearheaded the development of the cancer drug Gleevec.

STARRING INSTITUTES

Although Novartis is a major tenant, a significant amount of space at Biopolis will be taken up by five research institutes run by the Agency for Science, Technology and Research (A*STAR) — the Genome Institute of Singapore (GIS), the Institute of Bioengineering and Nanotechnology (IBN), the Bioinformatics Institute (BII), the Institute of Molecular and Cell Biology (IMCB) and the Bioprocessing Technology Institute. Philip Yeo, A*STAR's director, has spent the past few years securing his own stars to run these divisions. In March 2001, for example, a key member of the team took up his post as director of the GIS. The recruitment of Edison Liu, former head of clinical sciences at the US National Cancer Institute (NCI) in Bethesda, Maryland, sent out a clear signal that Singapore was serious about becoming internationally competitive.

Yeo's most recent coup was convincing Jackie Ying to leave the Massachusetts Institute of Technology (MIT) in Cambridge for Singapore to head the IBN. Ying, who was the youngest tenured professor ever at MIT, could not resist the chance to build something up from nothing. "I wanted an environment where there's a tremendous opportunity to grow," she says.

Ying has already started to build the institute up with researchers from the United States, Europe, Asia and Australia. The IBN has expanded from 20 to over 90 staff since she arrived this March. Eventually it will be home to 250 people, of whom 90 will be permanent staff with PhDs, and 90 a mix of postdocs and graduate students.

The GIS is also set for some significant growth. When the institute launched in 2001 it had a staff of 20, six of whom were scientists. It is now up to 180, with plans to grow to about 280 once it moves into Biopolis this year and to 350 by the end of 2005.

Lance Miller, who followed Liu from the NCI, took about six months to make up his mind to join his mentor. But after visiting Singapore, where he talked to

Mixed blessing

When Lisa Ng (pictured), a native Singaporean, started her postdoc at the Genome Institute of Singapore (GIS), she initially concentrated on hepatitis B. "And then SARS paid a visit," she says. Her background knowledge on the family of coronaviruses — of which SARS (severe acute respiratory syndrome) is a member — was a welcome coincidence. And at the GIS, her group helped to sequence five separate strains in the same time frame as it took a US group and a Canadian group to each deliver one. With that work the GIS researchers helped to show that SARS could mutate. Ng then used that information to develop a genetic screening test for SARS, which has since been licensed by Roche Diagnostics. Awareness of SARS in Singapore had already been high, thanks to posters in taxi cabs and lifts. Now awareness of Singapore science is similarly heightened, thanks to the work on the virus by the Biopolis-based researchers.



P.S.



The magnificent seven: when it is completed, the Biopolis complex will be home to around 1,500 scientists.

a number of scientists, he was convinced. The GIS' role in decoding the genome of the virus that causes severe acute respiratory syndrome (SARS) has since validated his decision (see 'Mixed blessing', opposite).

Miller, like many A*STAR scientists, is excited about the interconnectedness of the Biopolis institutes. Walkways join the GIS and three other A*Star institutes to one that is central — both literally and figuratively — to them all, the BII, which will provide informatics support to all of the surrounding institutes. Emphasizing the BII's importance, its building also houses the Biopolis cafeteria and lecture halls. "Everyone has to come to the BII for the seminars and the meals," says Gunaretnam Rajagopal, the institute's acting executive director.

The BII, like the other A*STAR institutes in Biopolis, is on course to grow when it moves in. It has 82 people now, including staff, postdocs and students, and will expand to 100 once it moves in this year, adding another 25 or so in two years' time.

Torsten Exner, a research scientist at the Bioprocessing Technology Institute, was another recruit who initially had doubts about moving to Singapore. "Singapore is not a nation that has a big history of research," Exner says. But while he was completing a postdoc at AstraZeneca in Sweden, Exner did an online search for scientists working in his research speciality and came up with a hit on Singapore. So he moved there in 2002 and has been pleased to be part of the institute's growth — its 70-member staff is set to double by 2005.

While other A*STAR institutes are trying to create their own history of research, the oldest, the IMCB, has

a head start and is sharpening its focus on developmental biology, structural biology, cancer biology and infectious diseases. Established 16 years ago, the IMCB has grown from 35 people when it opened, to about 400 scientists now. Moving into Biopolis will allow it to expand its number of students and postdocs by 50%. Hong Wanjin, the IMCB's deputy director, says that he is proud that the institute has trained over 100 PhD students since its inception. Its next step is to grow more Singapore talent. Of the 35 principal investigators now at the IMCB, only six hold Singapore passports.

CENTRALIZED RESERVATIONS

There are signs, too, that Singapore is overcoming its biggest challenge — growing independent science under an autocratic government. Alan Porter, a principal investigator at the IMCB, joined the institute at its inception. He was attracted by promises that he would be totally funded and independent — a promise that the government has largely kept, he adds. At various times, the government has pushed for more applied research at the IMCB, but has so far accepted that basic as well as applied research can generate intellectual property and contribute to the life-sciences push in Singapore, Porter notes.

For Singapore, the Biopolis infrastructure creates a good foundation. If the country creates a more open scientific culture, the government should have an easy time attracting world-class researchers to its world-class facility.

Paul Smaglik is editor of *Naturejobs*.

Web links

- Biopolis
- ♦ www.one-north.com/pages/lifeXchange/bio_intro.asp
- Agency for Science, Technology and Research
- ♦ www.a-star.gov.sg/astar/index.do
- Genome Institute of Singapore
- ♦ www.gis.a-star.edu.sg
- Institute of Bioengineering and Nanotechnology
- ♦ www.ibn.a-star.edu.sg
- Bioinformatics Institute
- ♦ www.bii.a-star.edu.sg
- Institute of Molecular and Cell Biology
- ♦ www.imcb.a-star.edu.sg
- Bioprocessing Technology Institute
- ♦ www.btc.nus.edu.sg
- Novartis Institute for Tropical Diseases
- ♦ www.nitd.novartis.com